Tax reforms and multi-dimensional screening*

Felix J. Bierbrauer Pierre C. Boyer Andreas Peichl Daniel Weishaar

March 26, 2026

Abstract

The key question in multi-dimensional screening problems is how prices, incentives, or marginal tax rates in one economic activity should vary with other activities. We develop a theory of tax reforms in a setting with multi-dimensional heterogeneity among taxpayers who make two economic decisions. Our leading application is the taxation of couples who choose an earnings level for each spouse. In our theoretical analysis, we characterize the conditions under which reforms of a given tax system yield Pareto- or welfare-improvements. In an empirical application to the US, we quantify the welfare implications of such reforms. We also prove an impossibility result: under assumptions common in the tax perturbation literature, the hypothesis that the given status quo tax is optimal leads to a contradiction. Thus, the perturbation approach cannot be used to characterize a fully optimal tax system.

Keywords: Tax reforms, Multi-dimensional screening, Taxation of Couples, Optimal taxation.

JEL classification: C72; D72; D82; H21.

*__Bierbrauer__: CMR - Center for Macroeconomic Research, University of Cologne, Germany (bierbrauer@wiso.uni-koeln.de); __Boyer__: CREST, École Polytechnique, Institut polytechnique de Paris, France (pierre.boyer@polytechnique.edu); __Peichl__: ifo Munich, LMU Munich, CESifo, IHS and IZA (peichl@econ.lmu.de); __Weishaar__: University of Cologne (d.weishaar@wiso.uni-koeln.de).
We benefitted from conversations with Piotr Dworczak, Antoine Ferey, Mike Golosov, Christian Hellwig, Patrick Lahr, Jean-Charles Rochet, Aleh Tsyvinski, and Nicolas Werquin, and from comments received at the 2025 GRAPE Summer School on Inequality-aware market design and the IFS-Stone-UCL workshop on Optimal Taxation and Policy Design. We are extremely grateful to Dan Feenberg for granting us access to the NBER TAXSIM model and providing feedback on our simulations. This paper uses TAXSIM v32 in combination with the Current Population Survey (CPS) data. The authors gratefully acknowledge support by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy – EXC2126/1 – 390838866 as well as via CRC TRR 190 (project number 280092119) and the Investissements d'Avenir (ANR-11-IDEX-0003/Labex Ecodec/ANR-11-LABX-0047), ANR-19-CE41-0011-01 and ANR-20-CE41-0013-01. Weishaar gratefully acknowledges support from the Joachim Herz Foundation and the Fritz Thyssen Foundation.

1 Introduction

At the heart of multi-dimensional screening problems is how the incentives in one economic activity should vary with the pursuit of other economic activities. A classic in public finance is whether the tax paid on capital income should also depend on labour income?[1] Related questions arise in industrial organization: How should the price paid for one good or service depend on a consumer's demand of other goods and services?[2] Problems at the intersection of public finance and market design are also related: Should the price paid for an apartment, a medical service or an electric vehicle depend on one's position in the income distribution?[3] This paper presents a new approach to multi-dimensional screening problems and then applies it to a challenge in the design of tax systems, the taxation of couples. For this application, a key question is how the marginal tax rate on the earnings of one spouse should change as the other spouse gets richer or poorer?[4] This is a question about *jointness*, or, more technically, about the cross-derivative of a tax function $(y_1, y_2) \mapsto T(y_1, y_2)$, where y_1 stands for primary and y_2 for secondary earnings.

Our focus is on reforms of a given status quo. Rather than asking *"What degree of jointness would be chosen by a welfare-maximizing planner?"*, we start from given a tax system with a predetermined degree of jointness and then analyze the consequences of reforming this system. In an empirical application, we look at the tax treatment of married couples in the US and analyze the consequences of reforms that would bring this system closer to individual taxation. To get there, we develop tools for an analysis of reforms in a setting with multi-dimensional screening that can be applied in the presence of two empirically relevant complications: First, the status quo tax system may involve discontinuities in marginal tax rates ("kinks") or tax payments ("notches"). Second, we allow for behavioral responses at the intensive and the extensive margin so that, at the individual level, behavioral responses to taxation may be discontinuous. For such a setting, our analysis clarifies conditions under which a given tax system admits Pareto- or welfare- improving reforms.

1. See Ferey, Lockwood, and Taubinsky (2024) for a recent contribution.
2. Rochet (2024) contains survey of the literature on multi-product pricing.
3. See, e.g., Akbarpour et al. (2025) and Ahlvik, Liski, and Mäkimattila (2024).
4. See Golosov and Krasikov (2025) for a recent contribution to the literature on the optimal taxation of couples.

As is well-known, the mechanism design approach and the analysis of tax perturbations are alternative routes that can be taken for the characterization of an optimal tax systems. The perturbation approach usually invokes the assumption that behavioral responses to taxation can be characterized by the derivative of a choice function.[5] We show that, in a setting with multi-dimensional screening, this assumption may conflict with optimality. Thus, one can either study improvements of a status quo using the perturbation method or one can study optimal taxes. One cannot study optimal taxes via the perturbation method.

In the methodological part of the analysis, we use monotone comparative statics results and envelope theorems to characterize the revenue and welfare implications of tax reforms (Milgrom and Shannon 1994; Milgrom and Segal 2002). Those are captured by functional derivatives[6] that provide answers to questions of the following sort: *"Suppose jointness is decreased for all couples where primary earnings are between 150,000 and 160,000 USD and secondary earnings between 80,000 and 87,000 USD. Under what conditions is this a self-financing tax cut? Under what conditions does it come with a loss of tax revenue? Under what conditions does it improve social welfare?"* We contrast these answers with those to an alternative type of question: *"Suppose that, for all couples with earnings in the same range, the marginal tax on secondary earnings is reduced, while the marginal tax on primary earnings does not change. Under what conditions is this a self-financing tax cut? Under what conditions does it come with a loss of tax revenue? Under what conditions does it improve social welfare?"* This clarifies the difference between (i) changes of marginal tax rates that are implied by a change of jointness and (ii) changes of marginal tax rates that leave jointness as it is. Our analysis shows that jointness is a genuine dimension of tax design: The effects of a reform that modifies jointness cannot be replicated by combinations of reforms that modify only one marginal tax rate at a time, and vice versa. A reduction of jointness can be self-financing or welfare-improving even when every stand-alone reduction of a marginal tax rate would cause a loss of tax revenue or social welfare.

5. This applies to the rich literature that characterizes optimal taxes via "sufficient statistics" formulas that invoke various elasticities.

6. A key step in the formal analysis is to show that these functional derivatives are well-defined even though there are discontinuities both in marginal tax rates and in the behavioral responses to taxation.

Limits of tax perturbations: An impossibility result. A necessary condition for the Pareto-efficiency or welfare-optimality of a tax system is that there neither is an improving change of a marginal tax rate nor an improving change of jointness. We show that, with welfare weights that are decreasing in income, it is generally not possible to satisfy both conditions simultaneously. For specificity, consider tax revenue, or, equivalently, Rawlsian welfare to be the objective. If the tax system is designed such that modifications of jointness cannot generate additional tax revenue, then there exists a revenue-increasing change to a marginal tax rate. Conversely, if marginal tax rates are set "optimally", then revenue can be increased by a change of jointness.[7]

The impossibility result is derived under assumptions that seem weak at first glance: There is a given tax function. It may have kinks and notches and taxpayers with different characteristics might be induced to bunch at the same economic outcome. Further assumptions imply that behavioral responses can be characterized by the derivative of a choice function. This impossibility result sheds a light on papers that use perturbation techniques to derive optimal tax formulas with general welfare weights in settings with multi-dimensional screening.[8] If the impossibility result applies, then a tax system satisfying these optimal tax formulas is not a fully optimal tax system. It may be "partially optimal" in the sense that is exhausts the scope for improvement in one class of tax reforms, while generating scope for improvement in another class of tax reforms. The impossibility result also has an implication for the empirical analysis of tax reforms. When it applies, sequences of reforms that either improve marginal tax rates or the system's degree of jointness do not converge to a fully optimal tax system. The scope for improvements is never exhausted. There always is something left to do.

We provide an interpretation of the impossibility result from a more general mechanism design perspective. By the taxation principle (Hammond 1979; Guesnerie 1995), studying taxes rather than direct revelation mechanisms is without

7. The impossibility does not hold for every tax system and every welfare function. For instance, we show that, in the absence of income effects, and with an unweighted utilitarian welfare function, a laissez-faire outcome can neither be improved by a change of marginal tax rates nor by a change of jointness.

8. Jacquet and Lehmann (2021b), Jacquet and Lehmann (2021a), Bergstrom and Dodds (2021), Bierbrauer, Boyer, and Hansen (2023), Ferey, Lockwood, and Taubinsky (2024), and Spiritus et al. (2025).

loss of generality, unless one restricts the set of tax functions. We do, however, impose restrictions which imply that almost all agents, whether bunching or not, have unique solutions to their choice problems. This rules out patterns of bunching with a positive mass of taxpayers who have multiple solutions to their choice problems.[9] Such allocations cannot be reached when taxpayers have unique optima. Thus, when the starting point is such a tax system, it can be improved. This is the force that leads to the impossibility result.[10]

Application: The Taxation of Couples. In an empirical application, we estimate the welfare and revenue consequences of tax reforms in the data. We study the taxation of married couples in the US. The US has a system with progressive and *joint taxation*, as do France or Germany.[11] The tax treatment of married couples is a contentious issue in these countries, and it has been for quite some time. Our analysis is motivated by the key question in these controversies: Is it desirable to move away from joint taxation?

To map theory to data, we impose structure on preferences and effort costs, allowing revenue functions to be expressed in terms of observable parameters—the income distribution, the tax schedule, and behavioral elasticities. These are estimated by combining CPS microdata and tabulated IRS tax records with tax rates from the NBER TAXSIM model and elasticity estimates from the literature.

We find that in parts of the income distribution, the marginal tax rates on secondary earnings are inefficiently high. Lowering them would be self-financing. Realizing these Pareto-improvements requires to condition on the "right" combinations of primary and secondary earnings. By contrast, cutting the tax rates on

9. Boerma, Tsyvinski, and Zimin (2022) show that such patterns do arise under an optimal mechanism. In that case, a tax perturbation of the optimal allocation would provoke behavioral responses that cannot be captured by the derivative of a choice function. Instead, left- and right-sided derivatives would be needed to distinguish between, say, the consequences of a local increase of jointness and of a local decrease of jointness. Optimality would then imply that none of these changes can have a positive welfare impact. Formally this would be captured by two inequalities as opposed to a first order condition.

10. All that being said, in our setting, incentive compatible allocations are continuous in the characteristics of taxpayers. Consequently, an optimal allocation can be well approximated by a piecewise linear one and the latter can be implemented with a piecewise linear tax function. Our analysis covers piecewise linear tax functions. Thus, while a fully optimal system cannot be reached when taxpayers have unique optimal choices, it can still be well approximated.

11. When there is progressive and joint taxation, there is a convex function $\bar{T}$ so that $T(y_1, y_2) = \bar{T}(y_1 + y_2)$. Consequently, the primary and the secondary earner face the same marginal tax rate and there is positive jointness, the cross-derivative of T is positive.

secondary earnings across the board would lead to a loss of revenue.

At the bottom of the income distribution, reductions of marginal tax rates are costly in terms of revenue. Through the lens of our model, such reforms reduce the unconditional transfers that are paid to those with no market-income, and increase the earnings incentives of those with some market income. We find that such reforms are rejected by a Rawlsian welfare function, but approved by large class of alternative welfare functions, including ones that concentrate weights at the bottom of the income distribution.

We also analyze the welfare implications of reforms that reduce the tax system's jointness. We find that reductions of jointness are not self-financing. They imply both lower marginal tax rates and a loss of tax revenue. They increase, however, the economy's overall surplus ("Total output minus effort costs"). Welfare also improves with "Feminist" weights that increase in the secondary earner's income share.

Finally, we characterize optimal tax rates at the top of the income distribution.[12] For our application, we find that the Pareto tail in the joint distribution of primary and secondary earnings is such that secondary earnings should be taxed at a lower rate than primary earnings.[13]

Related Literature. There is a rich literature on multi-dimensional screening problems, with applications in various domains. The focus is on solutions that maximize profits, economic surplus, or measure of social welfare, see Rochet (2024) for a survey. Recent contributions with a focus on optimal taxes are Boerma, Tsyvinski, and Zimin (2022), Ferey, Lockwood, and Taubinsky (2024), Golosov and Krasikov (2025), Akbarpour et al. (2025) and Ahlvik, Liski, and Mäkimattila (2024). We bring a reform perspective to this literature and identify conditions under which a given tax system can be improved. The question whether jointness is too high or too low is confronted directly by tracing the revenue and the welfare consequences of changes to the tax function's cross-derivative. To the best of our

12. We impose assumptions that parallel those in Diamond (1998) and which imply that an optimal tax system is linear at the top. At the top, behavioral responses to taxation can therefore be characterized by the derivative of a choice function.

13. Earlier work has emphasized differences in the behavioral responses of primary and secondary earners as a rationale for differential taxation. Our empirical findings imply that differential taxation is desirable even if the relevant elasticities are the same for primary and secondary earners.

knowledge, this is without precedent in the literature. In an empirical application, we use this approach to shed light on alternatives to the current tax treatment of married couples in the US.

Our analysis of tax perturbations differs from much of the previous literature in that we do not assume that the status quo tax system is well-behaved. There may be discontinuities both in tax payments and in marginal tax rates. We also allow for discontinuities in the behavioral responses to taxation; e.g. a tax reform may turn a dual earner couple into a single-earner couple or vice versa. Bergstrom and Dodds (2025) deal with related issues. Making sure that tax perturbations have an impact on tax revenue that can be captured by a functional derivative is an important step in our analysis and also in theirs. The approaches differ though and are complementary. We construct the functional derivatives explicitly from local tax reforms that can be analyzed using monotone comparative statics results. Bergstrom and Dodds (2025) start from the definition of a Gateaux-differential and then provide sufficient conditions for its existence. For Bergstrom and Dodds (2025) this step is instrumental for the identification of welfare weights that rationalize a given tax schedule.[14] Our focus is different: For us, understanding the revenue implications of tax perturbations is instrumental for drawing conclusions about desirable changes of marginal tax rates and of jointness.

There is a rich literature that studies the optimal taxation of couples.[15] Our work is most closely related to Kleven, Kreiner, and Saez (2009) and Golosov and Krasikov (2025) who also treat the taxation of couples as a problem of multi-dimensional screening. Kleven, Kreiner, and Saez (2009) focus on a setting in which a primary earner only makes intensive margin choices and a secondary earner only

14. They also use this *inverse tax approach* to draw conclusions about welfare-improving changes of tax burdens.

15. Its starting point is the seminal paper by Boskin and Sheshinski (1983). The subsequent literature has branched out in numerous ways. Non-linear taxes have been considered, labor supply responses at the intensive and the extensive margin have been taken into account, marital status has been treated as endogenous, see e.g. Schroyen (2003), Brett (2009), Kleven, Kreiner, and Saez (2009), Immervoll et al. (2011), Cremer, Lozachmeur, and Pestieau (2012), Gayle and Shephard (2019), Malkov (2021), Alves et al. (2024), Ales and Sleet (2022), and, most recently, Spiritus et al. (2025) and Golosov and Krasikov (2025). A complementary literature embeds the joint labor supply decisions of couples into quantitative dynamic models. It then traces out the tax and transfer system's implications for the labor market outcomes and the savings decisions of men and women or tax revenue, see, e.g., Guner, Kaygusuz, and Ventura (2012), Bick and Fuchs-Schündeln (2017), Borella, De Nardi, and Yang (2023) or Holter, Krüger, and Stepancuk (2023).

makes an extensive margin choice, with the implication that distribution of secondary earnings is degenerate, it has only two mass points. Golosov and Krasikov (2025) focus on spouses who both only make intensive margin choices, with the implication that there are no single-earner couples.[16] Both Kleven, Kreiner, and Saez (2009) and Golosov and Krasikov (2025) present results on optimal jointness, whereas we focus on reforms that alter a given tax system's jointness. Moreover, we use a richer model of economic behavior, one that is consistent both with a non-degenerate distribution of secondary earnings and with a positive mass of single-earner couples.

Our empirical analysis employs the NBER TAXSIM microsimulation model and CPS micro data, as well as tabulated tax return data from the IRS.[17] The microsimulation model uses rich data on individual characteristics so that we can elicit, at the level of an individual tax unit, the marginal tax rates and tax liability under the status quo and any alternative tax system. The evaluation of tax reforms rests on empirical estimates of the behavioral responses to taxation. A large literature exists estimating relevant elasticities and our assumptions on behavioral responses are informed by this literature.[18]

Outlook. Section 2 specifies the model on which our analysis is based. In Section 3 we show that the functional derivatives which capture the revenue and welfare implications of tax reforms are well-defined. Conditions under which a tax system can be improved are derived in Section 4. Section 5 contains the result that a tax system cannot both be optimal and satisfy "common assumptions" in the analysis of tax perturbations. Section 6 contains the empirical evaluation of the current tax treatment of married couples in the US. The last section contains concluding remarks.

16. In an extension, they show that fixed costs of working are without consequence for optimal taxes when there is random matching on the marriage market.

17. Our empirical approach builds on and extends work by Eissa, Kleven, and Kreiner (2008), Bargain et al. (2015) and Bierbrauer, Boyer, and Peichl (2021).

18. See, e.g., Blundell and Macurdy (1999), Blau and Kahn (2007), Eissa and Hoynes (2004), LaLumia (2008), Bargain, Orsini, and Peichl (2014).

2 The model

Agents engage in two economic activities that, respectively, yield incomes of y_1 and y_2. There is one consumption good. Disposable income is given by

$$C(y_1, y_2) = b \ + \ y_1 + y_2 - T(y_1, y_2) \ ,$$

where b is the transfer to an agent with no earnings, and $(y_1, y_2) \mapsto T(y_1, y_2)$ is a tax function. Thus, the total tax payment of an individual with earnings of (y_1, y_2) is given by $-b + T(y_1, y_2)$, and we let $T(0, 0) = 0$ without loss of generality. At points y at which the tax function is differentiable, we write $T_1(y)$ and $T_2(y)$ for the marginal tax rates on the earnings due to activities 1 and 2. We write $T_{12}(y)$ or $T_{21}(y)$ for the cross derivatives. Individuals choose y_1 and y_2 to maximize utility

$$U\Big(C(y_1, y_2), y_1, y_2, \omega_1, \omega_2\Big) - \varphi_1 \mathbb{1}(y_1 > 0) - \varphi_2 \mathbb{1}(y_2 > 0) \ . \tag{1}$$

The indicator functions take positive values when fixed costs reduce the utility an agent can reach.[19] The fixed costs associated with activity j are denoted by φ_j, $j \in \{1, 2\}$. We assume that the function U is additively separable so that

$$U(c, y_1, y_2, \omega_1, \omega_2) \quad = \quad u(c) - k_1(y_1, \omega_1) - k_2(y_2, \omega_2) \ .$$

Consumption utility u is continuously differentiable, increasing and concave. The functions k_1 and k_2 give, respectively, the variable costs of productive effort in activities and 1 and 2. We assume that these functions are continuously differentiable, increasing and convex in y_i, decreasing in ω_i, and have a negative cross derivative. The interpretation is that higher types have lower effort costs, both in absolute terms and at the margin.[20] An agent's type is summarized by $\theta = (\omega, \varphi)$, where $\omega = (\omega_1, \omega_2)$ and $\varphi = (\varphi_1, \varphi_2)$. The cross-section distribution of types is denoted by $\tilde{F}$. We assume that this distribution is atomless, has support on a compact domain

19. We consider fixed costs to have a behavioral responses to taxation both at the extensive and the intensive margin. There are applications – such as the taxation of couples – where extensive margin responses are empirically important.

20. With this setup, single crossing conditions hold: At any point in a y_1-c space, a higher value of ω_1 implies a flatter indifference curve. Indifference curves in the y_1-c space are, moreover, convex. Indifference curves in the y_2-c-space have the analogous properties.

$\Theta = \Omega \times \Phi$ and is continuously differentiable. Utility-maximizing choices generate a distribution of incomes F. The set $Y^*(\Theta)$, with generic entry $y^*(\theta) = (y_1^*(\theta), y_2^*(\theta))$, contains the agents' optimal choices. We denote the range of y^* by $\mathcal{Y} = \mathcal{Y}_1 \times \mathcal{Y}_2$, where, for $i = 1, 2$, $\mathcal{Y}_i = [0, \bar{y}_i]$.

Example 1: The taxation of couples. This is the application that we use in our quantitative application. Here, y_1 stands for primary earnings and y_2 for secondary earnings. Disposable income is treated as a public good for the household members. We also invoke an additively separable utility function that is quasi-linear in disposable income

$$U(c, y_1, y_2, \omega_1, \omega_2) \quad = \quad c - k_1(y_1, \omega_1) - k_2(y_2, \omega_2) \; .$$

The variable effort cost functions k_1 and k_2 are assumed to be iso-elastic

$$k_1(y_1, \omega_1) \;=\; \frac{1}{1 + \frac{1}{\varepsilon_1}} \left(\frac{y_1}{\omega_1} \right)^{1 + \frac{1}{\varepsilon_1}} \quad \text{and} \quad k_2(y_2, \omega_2) \;=\; \frac{1}{1 + \frac{1}{\varepsilon_2}} \left(\frac{y_2}{\omega_2} \right)^{1 + \frac{1}{\varepsilon_2}} \; , \tag{2}$$

where ε_1 and ε_2 are, respectively, the elasticities that govern the behavioral responses of primary and secondary earners. This specification is often used in the optimal taxation literature, e.g., Golosov and Krasikov (2025).

Example 2: Taxing labor and capital income. We switch to a setting with one source of income and two consumption goods. Consider the budget constraint

$$c_1 \quad = \quad b + y - c_2 - T(y, c_2) \; , \tag{3}$$

and the utility function $U(c_1, c_2, y, \omega, \beta) - \varphi_1 \mathbb{1}(y > 0) - \varphi_2 \mathbb{1}(c_2 > 0)$, with

$$U(c_1, c_2, y, \omega, \beta) \quad = \quad u_1(c_1) + u_2(c_2, \beta) - k_1(y_1, \omega_1) \; .$$

Here, u_1 gives the utility from consuming the first good and u_2 gives utility from consuming the second good. The function u_2 is increasing in both arguments and has a positive cross-derivative. Ferey, Lockwood, and Taubinsky (2024) use this setting to analyze the joint taxation of savings/ consumption in period 2 and income

y. The only difference to the setting in this paper is that, in Example 2, c_2 decreases the right hand side of the budget constraint and increases utility, whereas, in Example 1, y_2 increases the right-hand side of the budget constraint and decreases utility. The analysis that follows can be easily adapted to deal with this alternative specification.

3 Perturbing a wild tax system

This section sets the stage for the subsequent analysis. It shows that the revenue and welfare implications of tax reforms can be captured by functional derivatives even if the status quo tax system is wild (rather than smooth) and even if behavioral responses to taxation are discontinuous. Two assumptions are imposed in the following without further mention. Assumption 1 introduces the possibility of discontinuities in the tax function ("notches") or in marginal tax rates ("kinks"). Assumption 2 ensures that optimal choices are well-defined even if there are such discontinuities.

Assumption 1 *There is a partition $\mathcal{P} = \{P_1, \ldots, P_m\}$ of $\mathcal{Y}$ so that, for all j, T is continuously differentiable and convex over the interior of P_j. $\mathcal{P}$ is monotone in the following sense: for any y_{-i}, $(y_i, y_{-i}) \in P_k$ and $(y_i', y_{-i}) \in P_l$ imply $y_i < y_i'$ or $y_i > y_i'$.*

Under Assumption 1, kinks and notches may arise at the boundaries of the sets which form the partition $\mathcal{P}$. The convexity of T implies that the budget frontier is concave over the interior of the sets that form the partition.[21]

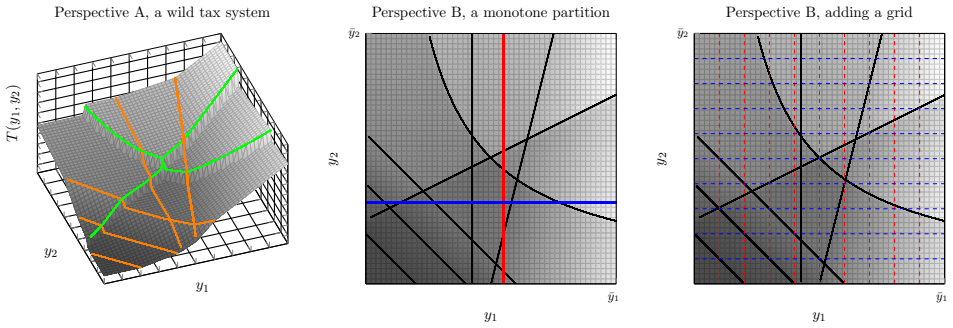
Assumption 2 $\forall y_{-i}$, $T(y_i, y_{-i}) = \min\left\{ \lim_{\epsilon\to 0^-} T(y_i+\epsilon, y_{-i}), \lim_{\epsilon\to 0^+} T(y_i+\epsilon, y_{-i}) \right\}$.

21. The assumption that T is convex is satisfied in our application. It is stronger than what is needed. For our theory to go through, we need that

$$y \quad \mapsto \quad u\big(b + y_1 + y_2 - T(y_1, y_2)\big) - k_1(y_1, \omega_1) - k_2(y_2, \omega_2) \ ,$$

is strictly concave over the interior of $P_j \in \mathcal{P}$, for those ω-types whose optimal choice lies in P_j . At the boundaries of the income space, we need the functions $y_1 \mapsto u(b+y_1-T(y_1,0))-k_1(y_1,\omega_1)$ and $y_2 \mapsto u(b+y_2-T(0,y_2))-k_2(y_1,\omega_2)$ to be strictly concave. This strict concavity is needed to ensure that intensive margin responses to changes of marginal tax rates can be captured by the derivative of a choice function; see also equations (50) - (53) in part D of the Appendix for an explicit computation of these derivatives. An inspection of these equations reveals that the derivatives are well defined if and only if the Hessian is strictly negative definite.

Figure 1: Adding a grid to a monotone partition



Note: The left figure shows a wild tax system with discontinuities in marginal taxes (kinks, orange lines) and taxes (notches, green lines). The middle figure shows a monotone partition with black lines reflecting kinks and notches in this tax system. If the partition is monotone the intersection of a horizontal or vertical line with a set P_j yields a compact interval. The figure on the right also shows the grid on which the definition of tax reforms is based. The intersection of the grid and the initial partition is a new monotone partition.

Assumption 2 specifies the tax function at the boundaries of the sets in $\mathcal{P}$. Whenever there is a notch so that, for some given y'_{-i}, there is y'_i at which the tax function, say, jumps upwards then the tax burden associated with an income of (y'_i, y'_{-i}) is the more attractive left-sided limit, i.e. $T(y_i, y_{-i}) = \lim_{\epsilon \to 0^-} T(y'_i + \epsilon, y'_{-i})$. As shown formally in the proof of Proposition 1 below, this assumption ensures that the choice problem in (1) actually has a solution.

3.1 Optimal behavior under a wild tax system

Proposition 1 below provides a characterization of optimal choices given taxes. An important insight for the subsequent analysis is that almost all types have a unique optimum. This simplifies the analysis of tax perturbations considerably. To be clear, and as illustrated in the proof of Proposition 1, there are types with multiple optima. These are either *cutoff types in* Ω who sit at kinks or notches or *cutoff types in* Φ who are indifferent between paying and not paying the fixed cost of an economic activity. By the arguments in Milgrom and Segal (2002), for those cutoff types, whenever the tax system is perturbed, left-sided and right-sided derivatives of choices and of realized payoffs do not coincide. By Proposition 1, these cutoff types have mass zero. An analysis that focuses on aggregate measures of tax revenue or social welfare can therefore proceed under the assumption that optimal choices and payoffs depend in a differentiable way on the parameters that define a tax

perturbation.

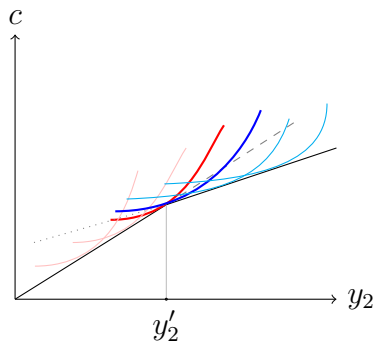
Cutoff types. To actually define *cutoff types in* Ω, consider the following subproblem: Let y_{-i} be fixed but arbitrary. For any P_k, define

$$v_k(\omega_i \mid y_{-i}) \quad := \quad \max_{y_i} \quad u(C(y_i, y_{-i})) - k_i(y_i, \omega_i) \quad s.t. \quad (y_i, y_{-i}) \in P_k \ . \tag{4}$$

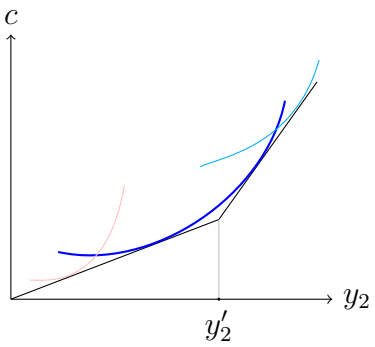
In the proof of Proposition 1, we verify that, given y_{-i}, there is a cutoff type $\hat{\omega}_i^{kl}$ separating those who choose $y_i \in P_k$ and those who choose $y_i \in P_l$, see Figure 2 for an illustration.

Figure 2: Kinks, Notches and cutoff types in Ω

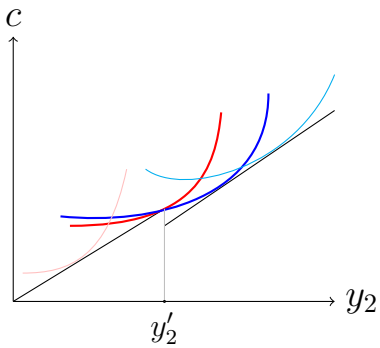
(a) A kink with MTR up



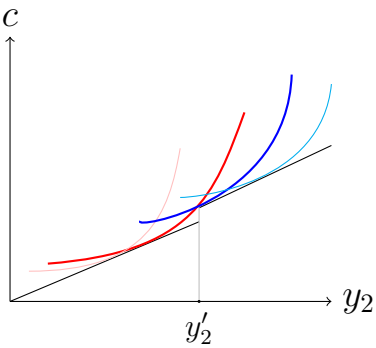
(b) A kink with MTR down



(c) A notch with T up



(d) A notch with T down



Note: The black line is the frontier of the budget set for a given level of y_1. The figure in the upper left shows an upward jump of the marginal tax rate (MTR) at y_2'. The bold blue and the bold red indifference curves are associated with cutoff types. Types in between bunch at y_2', whereas ω_2-types with an indifference curve flatter than the bold blue one, choose $y_2 > y_2'$, and ω_2-types with an indifference curve steeper than the red one choose $y_2 < y_2'$. The figure on the upper right shows a downward jump of the marginal tax rate at y_2'. Here, the bold blue indifference curve separates those ω_2-types who choose $y_2 > y_2'$, and those who choose $y_2 < y_2'$. The figure in the lower left (right) corner shows an upward (downward) jump of the tax burden T at y_2'. In both of these figures, the bold blue and the bold red indifference curves are associated with cutoff types.

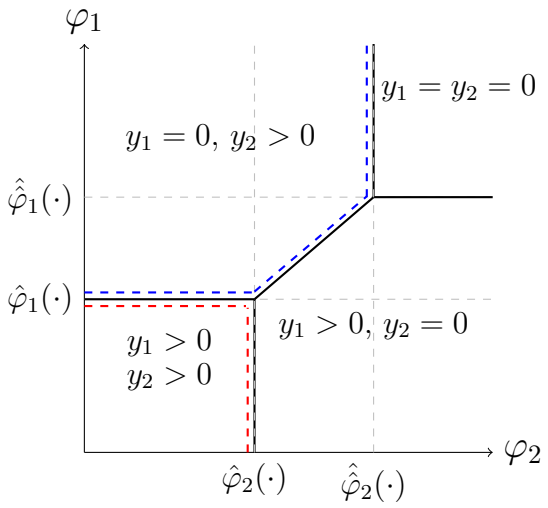
To define *cutoff types in* Φ, consider the subproblem to choose both y_1 and y_2

optimally conditional on paying the fixed costs associated with these activities,

$$v^{12}(\omega) \;:=\; \max_{(y_1,y_2)\in\mathcal{Y}} \; u(C(y_1,y_2)) - k_1(y_1,\omega_1) - k_2(y_2,\omega_2)\;. \tag{5}$$

Choosing y_1 optimally conditional on paying the fixed costs only for activity 1 yields a payoff that we denote by $v^1(\omega_1)$ and analogously for $v^2(\omega_2)$. We write $U^{12}(\theta) = v^{12}(\omega) - \varphi_1 - \varphi_2$, $U^1(\theta) = v^1(\omega_1) - \varphi_1$ and $U^2(\theta) = v^2(\omega_2) - \varphi_2$ for payoffs that include the losses from fixed costs; and $U^0 = v^0(b)$ for the payoff that is realized when there is no income at all. Now, it is optimal to choose both y_1 and y_2 positive when $U^{12}(\theta) \geq \max\{U^1(\theta), U^2(\theta), U^0\}$. Given $v^{12}(\omega)$, $v^1(\omega)$ and $v^2(\omega)$, this requires that φ_1 and φ_2 satisfy three linear inequalities. With an atomless type space, there is zero mass on those combinations of ω and φ where (at least) one of them holds with equality, see Figure 3.

Figure 3: Cutoff types in Φ for given ω



Note: For a given ω, the figure shows how earnings choices depend on fixed costs $\varphi = (\varphi_1, \varphi_2)$ when cost functions are iso-elastic and the tax function is progressive and has a positive cross-derivative, see part K of the Supplementary material for a proof. In the lower left, there are two sources of income. In the upper right , there are no earnings at all. In the upper left, there is only income from activity 2. In the lower right, there is only income from activity 1. The red curve indicates how cutoff types shift, when the payoff $v^{12}(\omega)$ goes down. The blue line indicates how the cutoff types shift when the payoff $v^2(\omega)$ goes down.

Proposition 1 *For all θ, the choice problem in (1) has a solution. The value function $\theta \mapsto V(\theta)$ is continuous. The allocation $\theta \mapsto y^*(\theta)$ can be implemented with a tax function that is continuous. For almost all θ, the choice problem in (1) has a unique solution. For almost all θ and all $i \in \{1,2\}$, if $y_i^*(\theta) > 0$, then either*

there is strict bunching

$$
\begin{aligned}
& u'\Big(C(y^*(\theta))\Big)\Big(1 - \lim_{\epsilon\to 0^+} T_{y_i}(y_i^*(\theta)+\epsilon, y_{-i}^*(\theta))\Big) \\
& \quad < \quad k_{y_i}(y_i^*(\theta),\omega_i) \\
& \qquad < \quad u'\Big(C(y^*(\theta))\Big)\Big(1 - \lim_{\epsilon\to 0^-} T_{y_i}(y_i^*(\theta)+\epsilon, y_{-i}^*(\theta))\Big) \ ,
\end{aligned} \tag{6}
$$

or a first order condition is satisfied,

$$
u'\Big(C(y^*(\theta))\Big)\Big(1 - T_{y_i}(y^*(\theta))\Big) \quad = \quad k_{y_i}(y_i^*(\theta),\omega_i) \ . \tag{7}
$$

3.2 Revenue implications of making it even wilder

A tax reform replaces the status quo tax function T by a new one, T^n, and it replaces the status quo transfer b by $b+\delta$. We write $T^n(y) \ = \ T(y) \ + \ s\,h(y)$ and refer to the function h as the reform direction and to the scalar s as the size of the reform. We focus on reform directions h with $h(0,0) = 0$. We define reform directions h with reference to their local implications. To prepare for this, consider a partition of $\mathcal{Y}_1 = [0,\bar{y}_1]$ into $J+1$ intervals of equal length ℓ_1. The intervals are indexed by $\mathcal{J} = \{0,1,\ldots,J\}$, with the understanding that the interval with index j ranges from $y_1^j = j\ell_1$ to $y_1^{j+1} = (j+1)\ell_1$. Likewise, we consider a partition of $\mathcal{Y}_2 = [0,\bar{y}_2]$ into $K+1$ intervals of equal length ℓ_2, with index set $\mathcal{K} = \{0,1,\ldots,K\}$. For $(j,k) \in \mathcal{J}\times\mathcal{K}$, let $y^{jk} = (y_1^j, y_2^k)$. The collection of points y^{jk} with $(j,k)\in\mathcal{J}\times\mathcal{K}$ generates a grid on $\mathcal{Y}$. Any y^{jk} is the lower left corner of a box B^{jk}. For all y in this box we modify the tax function so that

$$
c \quad \leq \quad C^{jk}(y) \quad = \quad b+\delta+y_1+y_2 - \Big(T(y_1,y_2) + s\ h^{jk}(y)\Big) \ ,
$$

where

$$
h^{jk}(y) \quad = \quad \mathbb{1}(y\in B^{jk})\Big(b^{jk} + \tau_1^{jk}(y_1 - y_1^{jk}) + \tau_2^{jk}(y_2 - y_2^{jk})\Big) \ .
$$

The overall reform direction h is then defined with reference to what's happening in the various boxes, $h = \sum_{\mathcal{J}\times\mathcal{K}} h^{jk}$. Any local direction h^{jk} is defined by three parameters: (i) the local intercept b^{jk} which reduces to the overall intercept $b+\delta$,

so that the intercept for agents making choices in box B^{jk} is $b + \delta - s\ b^{jk}$,[22] (ii) the change in marginal tax rates τ_1^{jk} that applies to earnings from activity 1 to the extent that those exceed y_1^{jk}, and (iii) the change in marginal tax rates τ_2^{jk} that applies to earnings from activity 2. Thus, for a given grid, the reform (s, h) can also be summarized by δ and the parameters which are collected in $(s\ b^{jk}, s\ \tau_1^{jk}, s\ \tau_2^{jk})$ for $jk\ \in\ \mathcal{J} \times \mathcal{K}$.

Such a reform makes the tax system even wilder. If the initial tax system was smooth, the reformed system is not. If the initial tax system had kinks and notches, the reformed one has even more. There are additional discontinuities which are associated with the edges of the boxes. Assumption 1 is still satisfied under T^n: There is a new finite partition $\mathcal{P}^n$ which is monotone if $\mathcal{P}$ is monotone. Moreover, T^n inherits continuous differentiability and convexity over the interior of the sets in $\mathcal{P}^n$ from T. In particular, since any h^{jk} is affine-linear, the initial and the new schedule have the same curvature at points where T and T^n are both differentiable. If we specify T^n at the boundaries of the sets in $\mathcal{P}^n$ in such a way that Assumption 2 holds then Proposition 1 continues to apply under T^n.

We write

$$\Delta\Big(\delta, (s\ b^{jk}, s\ \tau_1^{jk}, s\ \tau_2^{jk})\Big) \quad = \quad \mathbf{E}_\theta[T^n(y^*(\theta, s, h)] \ - \ R^0$$

for the change in overall revenue due to the reform, where $y^*(\theta, s, h)$ is the choice of an agent with type θ under the new schedule, and R^0 is overall tax revenue in the status quo. We limit attention to values of δ that do not exceed some (possibly very large) upper bound $\bar{\delta}$.[23] The reform is *budget balanced* if

$$\delta \quad = \quad \Delta\Big(\delta, (s\ b^{jk}, s\ \tau_1^{jk}, s\ \tau_2^{jk})\Big)\ , \tag{8}$$

i.e., if the change in tax revenue is absorbed by changes in the intercept of the consumption schedule. Proposition 2 below establishes that budget-balanced reforms are "well-behaved" and it provides a characterization of their revenue implications in a neighborhood of the status quo.

22. Note that b^{00} is not a free parameter, $h(0,0) = 0$ implies $b^{00} = 0$.

23. This enables us to make direct use of Brouwer's fixed point theorem in the proof of Proposition 2 below.

In the statement of the Proposition we use the following notation: We denote the partial derivative of $\Delta(\cdot)$ with respect to δ by $\Delta_\delta(\cdot)$. If this derivative is evaluated at the status quo, i.e. at $s = 0$, we write Δ^0_δ. The parameters y^{jk} and $\ell = (\ell_1, \ell_2)$ define the box B^{jk}. We write $\Delta_b(\cdot \mid y^{jk}, \ell)$ for the partial derivative of Δ with respect to sb^{jk}, $\Delta_1(\cdot \mid y^{jk}, \ell)$ for the partial derivative of Δ with respect to $s\tau_1^{jk}$ and analogously for $\Delta_2(\cdot \mid y^{jk}, \ell)$. If these derivatives are evaluated at the status quo, we write $\Delta^0_b(y^{jk}, \ell)$, $\Delta^0_1(y^{jk}, \ell)$ and $\Delta^0_2(y^{jk}, \ell)$.

Revenue implications of behavioral responses to taxation. Any change in the parameters that define h^{jk} provokes behavioral responses at the intensive and the extensive margin which feed back into tax revenue. To give a precise account of these effects, we adapt the notation introduced in the context of Proposition 1 to the reformed system. For an agent with type ω, let $Y^{12}(s, h, \omega)$ be the set of optimal choices under the reformed system conditional on paying both fixed costs. Denote the value of the objective reached at the optimum by $v^{12}(s, h, \omega)$. Define $Y^1(s, h, \omega)$, $v^1(s, h, \omega)$, $Y^2(s, h, \omega)$, and $v^2(s, h, \omega)$ analogously. Denote by $\Phi^{12}(s, h, \omega)$ the set of all φ so that it is optimal to pay both fixed costs and define $\Phi^1(s, h, \omega)$ and $\Phi^2(s, h, \omega)$ analogously. We denote by

$$\sigma^{12}(s, h, \omega) \quad = \quad \mathbf{E}_\varphi[\mathbb{1}(\varphi \in \Phi^{12}(s, h, \omega)) \mid \omega]$$

the fraction of type ω individuals who pay both fixed costs. We define $\sigma^1(s, h, \omega)$ and $\sigma^2(s, h, \omega)$ analogously.

For optimal choices in the status quo we will make use of the shorthands $y^0_d(\omega) := y^*_d(0, h, \omega)$, $y^0_{s1}(\omega) := y^*_{s1}(0, h, \omega)$ and $y^0_{s2}(\omega) := y^*_{s2}(0, h, \omega)$. Armed with this notation, we can use the law of iterated expectations and write the change in revenue as

$$\mathbf{E}_\theta[T^n(y^*(s, h, \theta)) - r^0(\omega)] \quad = \quad \mathbf{E}_\omega[r(s, h, \omega) - r^0(w)] \,,$$

where

$$\begin{aligned} r(s,h,\omega) &= \mathbf{E}_{\varphi}[T^n(y^*(s,h,\varphi,\omega))] \\ &= \sigma^{12}(s,h,\omega)\Big(T(y_d^*(s,h,\omega)) + s\ h(y_d^*(s,h,\omega))\Big) \\ &\quad +\sigma^{1}(s,h,\omega)\Big(T(y_{s1}^*(s,h,\omega),0) + s\ h(y_{s1}^*(s,h,\omega)\Big) \\ &\quad +\sigma^{2}(s,h,\omega)\Big(T(0,y_{s2}^*(s,h,\omega)) + s\ h(0,y_{s2}^*(s,h,\omega))\Big)\ . \end{aligned} \tag{9}$$

and $r^0(\omega) := r(0,h,\omega)$ is tax revenue in the status quo. The decomposition in (9) enables us to distinguish the reform's mechanical effect on tax revenue, captured by h, the revenue implications of behavioral responses at the intensive margin, associated with changes of $y_d^*(s,h,\omega)$, $y_{s1}^*(s,h,\omega)$ and $y_{s2}^*(s,h,\omega)$, and the extensive margin, associated with changes of $\sigma^{12}(s,h,\omega)$, $\sigma^1(s,h,\omega)$ and $\sigma^2(s,h,\omega)$.

Proposition 2 *There is a solution $R(s,h)$ to the fixed point equation in (8), it is unique if the tax function T is non-decreasing. The function $\Big(\delta,(s\ b^{jk},s\ \tau_1^{jk},s\ \tau_2^{jk})\Big) \mapsto R(s,h)$ is differentiable. The total derivative of $R(s,h)$ with respect to s, evaluated at $s=0$, is given by*

$$R_s(0,h) \ := \ \frac{1}{1-\Delta_\delta^0}\sum_{(j,k)}\Big(b^{jk}\Delta_b^0(y^{jk},\ell)+\tau_1^{jk}\Delta_1^0(y^{jk},\ell)+\tau_2^{jk}\Delta_2^0(y^{jk},\ell)\Big)\ .$$

For $x \in \{sb^{jk}, s\tau_1^{jk}, s\tau_2^{jk}\}$,

$$\Delta_x^0(y^{jk},\ell) \quad = \quad \mathbf{E}_\omega\Big[\mathbb{1}\left(y_d^0(\omega)\in B^{jk}\right)\ \Big(\beta_{I,x}^0(\omega)+\beta_{E,x}^0(\omega)+\mu_x^0(\omega)\Big)\Big]\ , \tag{10}$$

where $\beta_{I,x}^0(\omega)$, $\beta_{E,x}^0(\omega)$ and $\mu_x^0(\omega)$ capture, respectively, the revenue implications of intensive margin responses (changes of y_d^*), extensive margin responses (changes of σ^{12}, σ^1 and σ^2) and the mechanical effect from changes of the tax function.

By Proposition 2, the functional derivative of tax revenue in direction h is well-defined. Moreover, the reform's total impact on tax revenue is a sum of the local effects that are associated with the various boxes. The effects from a given box B^{jk} are a linear combination of what is due to the change of the local intercept and what is due to the change of the marginal tax rates on the earnings from activities 1 and 2, respectively. Generally, the sorting of taxpayers across boxes is affected by the reform. More formally, the cutoff types discussed above depend both on s and h.

By Equation (10), this change does not affect the revenue implications of reforms that are "small" and have marginal effects evaluated at the status quo, i.e. at $s = 0$. We refer to this as the "*What happens in the box, stays in the box*"-property: a local change of the tax system has no effect on the revenue generated elsewhere. In part H.1 of the Supplementary material we provide a more detailed characterization of the terms that appear in Equation (10).[24]

3.3 Welfare implications of tax reforms

It takes only a small step to get from a reform direction's implication for tax revenue to its implications for social welfare. Denote by $V(s, h, \theta)$ the indirect utility realized by an agent with type θ under the perturbed tax schedule. It follows from Corollary 4 in Milgrom and Segal (2002) that for all types with a unique optimum in the status quo, we have that

$$V_s(0, h, \theta) \quad = \quad u'(c^0(\theta))\Big(R_s(0, h) - h(y^0(\theta))\Big) \ ,$$

where $y^0(\theta) = (y_1^0(\theta), y_2^0(\theta))$ are earnings and $c^0(\theta)$ is consumption in the status quo.[25] We say that reform direction h is Pareto-improving if, for all θ,

$$R_s(0, h) - h(y^0(\theta)) \quad \geq \quad 0 \ ,$$

with a strict inequality for some θ. Given a specification of welfare weights $\theta \mapsto g(\theta)$, we say that reform direction h is welfare-improving if

$$W_s(0, h) \quad = \quad \mathbf{E}_\Theta \Big[g(\theta)\, u'(c^0(\theta)))\Big(R_s(0, h) - h(y^0(\theta))\Big)\Big] \quad > \quad 0 \ . \tag{11}$$

24. This part is relegated to the Supplement because it is mechanical (computing derivatives and collecting terms). One insight there is that strict bunchers show no behavioral responses at the intensive margin. More specifically, if they are strict bunchers both with respect to y_1 and with respect to y_2, then $\beta^0_{I,1}(\omega) = \beta^0_{I,2}(\omega) = 0$ for all x. If they are strict bunchers, say, only with respect to y_1 but not with respect to y_2, then $\beta^0_{I,1}(\omega) = 0$.

25. For types with multiple solutions to their utility-maximization problem, the right-side derivative of $V(s, h, \theta)$ with respect to s may not coincide with the left-sided derivative, hence, for these types, one has to explicitly distinguish the cases $s \to 0^+$ and $s \to 0^-$ to determine the impact of the perturbation on their payoff. For the purposes of this paper, there is no need to go into this any further. These types have mass 0 by Proposition 1; hence for measures of aggregate welfare all this is without consequence. Moreover, payoffs are continuous by Proposition 1. Thus, if a reform is Pareto-improving amongst agents with a unique optimum, then it is Pareto-improving in the set of all agents.

3.4 From types to incomes

Proposition 2 goes over the various boxes. In each box B^{jk}, there are changes of the tax function. Proposition 2 collects the revenue implications of all these changes by looking separately, for each box, at the types making choices in the box. A change of parameter x has revenue implications given by (10). The right hand side of (10) can also be written as

$$\int_{y\in B^{jk}} \Big(\bar{\beta}^0_{I,x}(y) + \bar{\beta}^0_{E,x}(y) + \bar{\mu}^0_x(y)\Big) f(y)dy \ , \tag{12}$$

where $x \in \{b, 1, 2\}$ indicates whether effects are due to the change of the local intercept, or the tax rates on activities 1 and 2, and the bar-symbol is used because the construction involves an averaging over ω-types who choose the same y. Finally, $y \mapsto f(y)$ is a density representing the status quo income distribution.[26]

Upon normalizing welfare weights such that, on average, they are equal to 1, and upon defining $\bar{g}(y') := \mathbf{E}_\theta[g(\theta)\ u'(c^0(\theta) \mid y^0(\theta) = y']$, we can write the welfare implications of a reform direction h also as

$$W_s(0,h) \quad = \quad R_s(0,h) - \mathbf{E}_y\left[\bar{g}(y)\, h(y)\right] \quad > \quad 0 \ ,$$

where expectations are computed using the empirically observed distribution of incomes.

3.5 Revenue implications of continuous reform directions

Proposition 3 below provide a characterization of $R_s(0,h)$ for a continuous reform directions $y \mapsto h(y)$ that will be used in the subsequent sections.[27] To prove Proposition 3, we first approximate the revenue implications of h using a sequence of piecewise linear functions $\{h^n\}$. The revenue implications of the latter come from Proposition 2. We then make the grid finer and finer and in the limit obtain

26. Part H.1 of the Supplementary Material gives a formal definition of the terms on the right-hand side of (12) both for income levels y in the atomless part of the income distribution and for mass points.

27. Note that Proposition 2 does not require continuity of the reform direction h. With $h = \sum_{\mathcal{J}\times\mathcal{K}} h^{jk}$ continuity holds within boxes but not across boxes.

$R_s(0,h) = \lim_{n\to\infty} R_s(0,h^n)$ as a Riemann integral.[28]

Proposition 3 *Let $y \mapsto h(y)$ be a continuous function with $h(0,0) = 0$. Then*

$$(1-\Delta_\delta^0)\, R_s(0,h) \;=\; \int_{\mathcal{Y}} \Big(h(y)\, \bar{\mathcal{R}}_b(y) + h_1(y)\, \bar{\mathcal{R}}_1(y) + h_2(y)\, \bar{\mathcal{R}}_2(y) \Big) dy \;, \quad (13)$$

where $\Delta_\delta^0 = \mathbf{E}_y\left[\bar{\beta}_{E,b}^0(y) + \bar{\beta}_{I,b}^0(y)\right]$, $\bar{\mathcal{R}}_b(y) = \big(\bar{\beta}_{I,b}^0(y) + \bar{\beta}_{E,b}^0(y) + \bar{\mu}_b^0(y)\big)\, f(y)$ and

$$\bar{\mathcal{R}}_1(y) = f(y)\, \bar{\beta}_{I,1}^0(y) \;, \quad \bar{\mathcal{R}}_2(y) = f(y)\, \bar{\beta}_{I,2}^0(y) \;, \quad (14)$$

If $u(c) = c$, then for all y, $\bar{\beta}_{I,b}^0(y) = 0$ and $\mathbf{E}_y\left[\bar{\beta}_{E,b}^0(y)\right] = 0$, and, hence, $\Delta_\delta^0 = 0$.

For $x \in \{1,2\}$, how do we get from the collection of terms in Equation (12) to the more compact expressions in (14)? Some of the terms in (12) vanish as the grid gets finer and finer and the lengths of the boxes go to zero. For instance, we have that $\mu_1^0(\omega) = \sigma^{12}(\cdot)\big(y_{d1}^*(\cdot) - y_1^{jk}\big)$ and, since $0 \leq y_{d1}^*(\cdot) - y_1^{jk} \leq \ell_1$, this term vanishes as $\ell_1 \to 0$. The same logic implies that $\mu_2^0(\omega)$, $\beta_{E,1}^0(\omega)$ and $\beta_{E,2}^0(\omega)$ vanish as $\ell_2 \to 0$. In the proof of Proposition 3 we also verify that, in the absence of income effects, a local change of the tax burden b does not come with behavioral responses at the intensive margin, $\bar{\beta}_{I,b}^0(y) = 0$. Moreover, a change of the intercept δ, akin to a global change of the tax burden, does not come with behavioral responses at the extensive margin. This implies that $\Delta_\delta^0 = 0$ when $u(c) = c$.

A special case. For the case $u(c) = c$, we have that

$$\bar{\mathcal{R}}_1(y) = \Big[T_1(y) y_{d1,1}^0(\cdot) + T_2(y) y_{d2,1}^0(\cdot) \Big] f(y) \;, \quad (15)$$

where $y_{d1,1}^0(\cdot)$ and $y_{d2,1}^0(\cdot)$ are derivatives that capture intensive margin responses to a local change of τ_1.[29] Moreover,

$$\bar{\mathcal{R}}_b(y) = \Big(1 - L_E(y)\Big) f(y) \quad (16)$$

28. The arguments used to prove Proposition 3 also apply to reform directions h that are continuous only over subsets of the space of incomes and equal to zero elsewhere (see Figure 4 for an example).

29. Equation (15) follows from squaring the derivations in part G.2 of the supplementary material with the implications of $u(c) = c$, see Proposition 3.

where $L_E(y)$ is a revenue loss from extensive margin responses. In the absence of fixed costs, this expression simplifies:

$$\bar{\mathcal{R}}_b(y) = f(y) \ . \tag{17}$$

In our empirical application, we use these expressions. We assume, moreover, that the status quo tax system is piecewise linear, as is the case in the US, and that the effort cost functions take the iso-elastic form in (2). We then obtain,[30]

$$\begin{aligned} \bar{\mathcal{R}}_1(y_1, y_2) &= -\varepsilon_1 \ y_1 \ \frac{T_1(y_1, y_2)}{1 - T_1(y_1, y_2)} f(y_1, y_2) \ , \\ \bar{\mathcal{R}}_2(y_1, y_2) &= -\varepsilon_2 \ y_2 \ \frac{T_2(y_1, y_2)}{1 - T_2(y_1, y_2)} f(y_1, y_2). \end{aligned} \tag{18}$$

These formulas apply both in the interior of the income space and at its boundary, with one qualification: When $y_1 > 0$ and $y_2 > 0$, then $L_E(y_1, y_2)$ captures the fiscal externalities that are associated with agents giving up one activity (see the red dashed line in Figure 3). When, say, $y_1 = 0$, and the tax burden for people with income only from activity 2 goes up, then $L_E(y_1, y_2)$ captures the fiscal externalities that are associated with the fact that some of these agents now also give up on activity 2 and end up with $y_1 = y_2 = 0$, others specialize on activity 1 and have $y_1 > 0$ and $y_2 = 0$ after the reform, yet another group moves into the category with $y_1 > 0$ and $y_2 > 0$ (see the blue dashed line in Figure 3). In this case, the total effect captured by $L_E(y_1, y_2)$ may actually be negative.[31]

4 Evaluating directions for reform

4.1 Local changes of one marginal tax rate

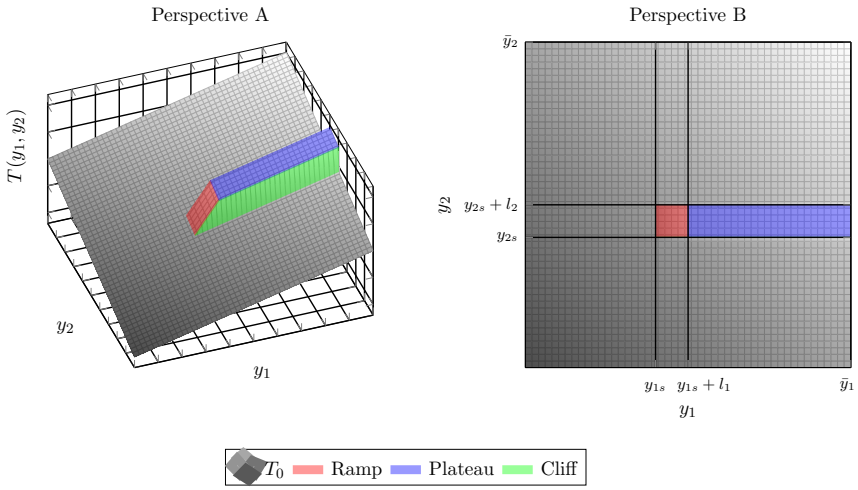
Fix two brackets $B_1 = [y_{1s}, y_{1s}+\ell_1]$ and $B_2 = [y_{2s}, y_{2s}+\ell_2]$. Define a reform direction h^j so that for $y \in B := B_1 \times B_2$, $h^j(y) = \tau_1^j(y_1 - y_{1s})$ which implies that $h_1^j(y) = \tau_1^j$ and $h_2^j(y) = 0$. For $y_1 \geq y_{1s} + \ell_1$ and $y_2 \in B_2$, let $h^j(y) = \ell_1 \ \tau_1^j$. Otherwise let

30. The computations are in part K of the Supplementary material.

31. In part K of the Supplementary material, we show that $L_E(y)$ can be written as a sum of extensive margin elasticities that capture the percentage share of people who leave one of the areas in Figure 3 and enter into another one, in response to a one-percent change of the tax savings that are associated with such a switch of categories.

$h^j(y) = 0$. Figure 4 shows such a function. It consists of a ramp and a plateau. At the edges of the ramp and the plateau, there is a cliff.[32]

Figure 4: Changing one marginal tax rate



Note: This figure illustrates the implications of changing marginal tax rates in one dimension. The marginal tax rate is changed for activity 1 for agents with $y_1 \in B_1$ and $y_2 \in B_2$ (red ramp). The change in marginal tax rates leads to higher taxes for agents with $y_1 > y_{1s} + \ell_1$ and $y_2 \in B_2$ (blue plateau) and generates a discrete change in marginal tax rates at $y_{1s} + \ell_1$ and $y_{2s} + \ell_2$ (green cliff).

Using Proposition 3 and an integration by parts,[33] yields a characterization of the reform's revenue impact

$$(1 - \Delta_\delta^0)\, R_s(0, h^j) \;=\; \tau_1^j \int_{y_{2s}}^{y_{2s}+\ell_2} \int_{y_{1s}}^{y_{1s}+\ell_1} \mathbf{R}_1(y_1 \mid y_2)\, dy_1\, dy_2. \tag{19}$$

The function $y \mapsto \mathbf{R}_1(y_1 \mid y_2)$ is obtained as follows: In the derivation of Equation (19) we use an integration by parts to replace

$$\int_{y_{2s}}^{y_{2s}+\ell_2} \int_{y_{1s}}^{\bar{y}_1} \Big(h^j(y)\, \bar{\mathcal{R}}_b(y) + h_1^j(y)\, \bar{\mathcal{R}}_1(y) \Big) dy_1\, dy_2$$

32. Thus, a sizable reform in direction h^j generates dominated regions in the income space; e.g. with $\tau_1^j > 0$, jumping off the cliff makes agents better off so that the edges of the ramp and the plateau are dominated. For a small reform that is evaluated at the status quo, i.e. at $s = 0$, this does not create complications. By Proposition 2, we can evaluate revenue implications assuming that people respond to the reform while being sorted across the income space as in the status quo.

33. The computations are documented in Section I of the Supplementary material.

by

$$\int_{y_{2s}}^{y_{2s}+\ell_2} \int_{y_{1s}}^{y_{1s}+\ell_1} h_1^j(y_1, y_2)\Big(\bar{\mathcal{R}}_1(y_1, y_2) + \textstyle\int_{y_1}^{\bar{y}_1} \bar{\mathcal{R}}_b(s_1, y_2) ds_1\Big) dy_1\, dy_2 \quad .$$

This captures that, given a fixed $y_2 \in B_2$, a change of the marginal tax rate at y_1 is associated with a larger tax burden for all $y_1' > y_1$. At y_1, there is a behavioral response due to the change in marginal tax rates as captured by $\bar{\mathcal{R}}_1(y_1, y_2)$. For every $y_1' > y_1$, there is a mechanical change of the tax burden which also provokes behavioral responses. All this is captured by $\int_{y_1}^{\bar{y}_1} \bar{\mathcal{R}}_b(s_1, y_2) ds_1$. The function $y \mapsto \mathbf{R}_1(y_1 \mid y_2)$ with

$$\mathbf{R}_1(y_1 \mid y_2) \quad := \quad \bar{\mathcal{R}}_1(y_1, y_2) + \textstyle\int_{y_1}^{\bar{y}_1} \bar{\mathcal{R}}_b(s_1, y_2) ds_1 \tag{20}$$

gives the total revenue effect associated with a change of the marginal tax rate at a given $y_1 \in B_1$, for a given $y_2 \in B_2$. In getting to the reform's revenue impact we "sum" these terms and go over all $y_1 \in B_1$ and all $y_2 \in B_2$. For brevity, we refer to $\mathbf{R}_1$ in the following as a *conditional revenue function*. We get analogous expressions when looking at a reforms where the marginal tax rate on activity 2 changes, while the marginal tax rate on activity 1 does not change. This yields the conditional revenue function $(y_1, y_2) \mapsto \mathbf{R}_2(y_2 \mid y_1)$.

Self-financing tax cuts. A necessary condition for the Pareto-efficiency of a tax system is that there are no self financing tax cuts. Pareto-efficiency therefore requires that, for all y, $\mathbf{R}_1(y_1 \mid y_2) \geq 0$ and $\mathbf{R}_2(y_2 \mid y_1) \geq 0$. In our application, we use these conditions to identify the parts of the US income distribution where marginal tax rates on secondary earnings are inefficiently high.

Welfare-implications. Using $W_s(0, h) = R_s(0, h) - \mathbf{E}_{\mathcal{Y}}\left[\bar{g}(y)\, h(y)\right]$, our analysis of welfare implications closely mirrors the analysis of revenue implications, except that the welfare losses, say, from increases of the tax burden show up as an extra term. Specifically, $\mathbf{E}_{\mathcal{Y}}\left[\bar{g}(y)\, h(y)\right]$ can, after an integration by parts, be written as

$$\tau_1^j \int_{y_{2s}}^{y_{2s}+\ell_2} \int_{y_{1s}}^{y_{1s}+\ell_1} \left(1 - F_1(y_1 \mid y_2)\right) \mathcal{G}(y_1 \mid y_2) f_2(y_2) dy_1 dy_2,$$

where $\mathcal{G}(y_1 \mid y_2) = \mathbf{E}_{\mathcal{Y}}[\bar{g}(s_1, y_2) \mid s_1 \geq y_1]$ and the conditional *cdf* $F_1(\cdot \mid y_2)$ as well as the marginal distribution f_2 are derived from the mixed income density f. Upon defining

$$\mathbf{W}_1(y_1 \mid y_2) \;:=\; \tfrac{\mathbf{R}_1(y_1|y_2)}{1-\Delta_\delta^0} - (1 - F_1(y_1 \mid y_2))\, \mathcal{G}(y_1 \mid y_2) f_2(y_2)$$

we can state that

$$W_s(0, h^j) \;=\; \tau_1^j \int_{y_{2s}}^{y_{2s}+\ell_2} \int_{y_{1s}}^{y_{1s}+\ell_1} \mathbf{W}_1(y_1 \mid y_2)\, dy_1\, dy_2. \tag{21}$$

The *conditional welfare function* $y \mapsto \mathbf{W}_1(y_1 \mid y_2)$ is a modified version of the conditional revenue function. It incorporates changes of individual welfare that result from changes of the tax burden and applies welfare weights to these changes. Equation (21) is the welfare-analogue to Equation (19).

Optimal marginal tax rates. A necessary condition for welfare-optimality is that, for all y, $\mathbf{W}_1(y_1 \mid y_2) = 0$ and $\mathbf{W}_2(y_2 \mid y_1) = 0$. Using these conditions, the well-known formula for optimal top tax rates due to Diamond (1998) can be extended to the given setup. We impose additional assumptions which mirror those in Diamond (1998) and extend them to a setting where taxpayers have two sources of income.

Assumption 3 *Let $u(c) = c$. In addition, suppose that there is a range of incomes $Y' = Y_1' \times Y_2'$ over which the following properties all hold true: (i) There are numbers $\bar{g}_1$ and $\bar{g}_2$ so that $\mathcal{G}(y_1 \mid y_2) = \bar{g}_1$ and $\mathcal{G}(y_2 \mid y_1) = \bar{g}_2$, for all $(y_1, y_2) \in Y'$. (ii) There are numbers $\bar{\alpha}_1$ and $\bar{\alpha}_2$ so that*

$$\frac{f_{y_1}(y_1 \mid y_2)\, y_1}{1 - F_{y_1}(y_1 \mid y_2)} \;=\; \bar{\alpha}_1 \quad \textit{and} \quad \frac{f_{y_2}(y_2 \mid y_1)\, y_2}{1 - F_{y_2}(y_2 \mid y_1)} \;=\; \bar{\alpha}_2\,,$$

for all $(y_1, y_2) \in Y'$. (iii) Choices satisfy first order conditions and there are no behavioral responses at the extensive margin. (iv) Effort cost functions are iso-elastic.

These assumptions do not formally refer to the top of the income distribution. Empirically, however, they are most plausible near the top. For Y' near the top, part

(i) says that welfare weights do not discriminate between "the rich" and "the super rich". They may discriminate, however, between the sources of income which is the case when $\bar{g}_1 \neq \bar{g}_2$. Part (ii) says that, in the relevant income range, conditional income distributions are Pareto-distributions, and moreover that the Pareto-parameter for the conditional income distribution of y_i is, locally, independent of y_{-i}.

Proposition 4 *Suppose Assumption 3 holds. Define t_i^* implicitly by*

$$\frac{t_i^*}{1-t_i^*} \;=\; \Big(1-\bar{g}_i\Big)\,\frac{1}{\bar{\alpha}_i}\,\frac{1}{\varepsilon_i}.$$

If $T_1(y) = t_1^*$ and $T_2(y) = t_2^*$ for $y \in Y'$ then $\mathbf{W}_1(y_1 \mid y_2) = \mathbf{W}_2(y_2 \mid y_1) = 0$.

In our application, we use this characterization for a quantitative analysis of "optimal" top tax rates on primary and secondary earnings.

Welfare Trade-offs in the reduction of marginal tax rates. Suppose, for simplicity, that $u(c) = c$. Then, we can write $\mathbf{W}_2(y_2 \mid y_1)$ as

$$\mathcal{R}_2(y_2 \mid y_1) + \big(1 - F_2(y_2 \mid y_1)\big)\big(1 - L_E(y_2 \mid y_1) - \mathcal{G}(y_2 \mid y_1)\big) f_1(y_1)\;, \tag{22}$$

where $L_E(y_2 \mid y_1)$ captures revenue losses due to extensive margin responses. Now suppose that $\mathcal{R}_2(y_2 \mid y_1) < 0$ so that behavioral responses at the intensive margin lead to a loss of tax revenue[34] and that the status quo is an efficient tax system in the sense that $\mathbf{R}_2(y_2 \mid y_1) > 0$. Then $\mathcal{G}(y_2 \mid y_1) > 1$ implies

$$\mathbf{W}_2(y_2 \mid y_1) \quad < \quad 0 \quad < \quad \mathbf{R}_2(y_s \mid y_1)\,.$$

The property $\mathcal{G}(y_2 \mid y_1) > 1$ is plausible at the bottom of the income distribution.[35] When $\mathbf{W}(y_{2s} \mid y_{1s}) < 0$, and $0 < \mathbf{R}(y_{2s} \mid y_{1s})$, then the lowering of marginal tax rates on y_2 has two effects: First, it reduces the intercept of the consumption

34. One can show that $\mathcal{R}_2(y_2 \mid y_1) < 0$ holds if the status quo has progressive and joint taxation, i.e. with $T(y_1, y_2) = \bar{T}(y_1 + y_2)$, for some increasing and convex function $\bar{T}$.

35. In the context of couples taxation, it says: Consider, say, y_1 from the bottom decile of the marginal distribution of primary earnings. Compute an average welfare weights for all couples with that level of primary earnings. If y_1 is low, then this weight exceeds the average weight for the population at large (which has been normalized to 1).

schedule, and therefore is harmful to those with zero earnings both before and after the reform. A Rawlsian welfare function therefore rejects the reform. Second, it makes the budget line steeper in the sense that disposable income rises more strongly in y_2. This resembles introducing or strengthening elements of workfare in the tax and transfer system: Less unconditional transfers and more high-powered earnings incentives. It is remarkable that such reforms are desirable for a large class of social welfare functions, in fact for all welfare functions with $\mathcal{G}(y_2 \mid y_1) > 1$ at the bottom. In our application, we will verify $\mathbf{W}_2(y_2 \mid y_1) < 0 < \mathbf{R}_2(y_s \mid y_1)$ holds for "near-Rawlsian" welfare functions that concentrate weights in the bottom decile of the income distribution.

4.2 Local changes of jointness

Corollary 1 below provides an alternative characterization of $R_s(0,h)$ which exploits that a reform direction h can be characterized via its effects on marginal tax rates on the boundary of the income space and via its effects on jointness in the interior of the income space:[36]

$$h(y) \;=\; \int_0^{y_1} h_1(s_1,0)\,ds_1 \;+\; \int_0^{y_2} h_2(0,s_2)\,ds_2 \;+\; \int_0^{y_2}\int_0^{y_1} h_{12}(s_1,s_2)\,ds_1\,ds_2\,. \tag{23}$$

This observation together with a repeated use of integration by parts yields the alternative characterization of $R_s(0,h)$ and $W_s(0,h)$ in Corollary 1 below.

Corollary 1 *Let $y \mapsto h(y)$ be a continuous function with $h(0,0) = 0$. Then*

$$\begin{aligned}(1-\Delta_\delta^0)\,R_s(0,h) \;=\; &\int_0^{\bar{y}_1} h_1(y_1,0)\;\bar{\mathbf{R}}_1(y_1)dy_1 \quad + \quad \int_0^{\bar{y}_2} h_2(0,y_2)\;\bar{\mathbf{R}}_2(y_2)\;dy_2 \\ &+ \int_0^{\bar{y}_1}\int_0^{\bar{y}_2} h_{12}(y)\;\mathbf{R}^J(y)\;dy_2\;dy_1,\end{aligned} \tag{24}$$

36. The Corollary is obtained via a repeated use of integration by parts. Section J of the Supplementary material documents the requisite computations.

and

$$\begin{aligned} W_s(0,h) &= \int_0^{\bar{y}_1} h_1(y_1,0)\ \bar{\mathbf{W}}_1(y_1)dy_1 + \int_0^{\bar{y}_2} h_2(0,y_2)\ \bar{\mathbf{W}}_2(y_2)dy_2 \\ &\quad + \int_0^{\bar{y}_1}\int_0^{\bar{y}_2} h_{12}(y)\ \mathbf{W}^J(y)\ dy_2\ dy_1\ . \end{aligned} \tag{25}$$

In the characterization of $R_s(0,h)$, the changes of marginal tax rates at the boundary are weighted by an integral over conditional revenue functions. Specifically, the changes of marginal tax rates on the earnings from activity 2 are weighted by

$$y_2 \mapsto \bar{\mathbf{R}}_2(y_2) := \int_0^{\bar{y}_1} \mathbf{R}_2(y_2 \mid y_1)dy_1\ .$$

The changes of marginal tax rates on the earnings from activity 1 are weighted by $y_1 \mapsto \bar{\mathbf{R}}_1(y_1)$, analogously defined. This reflects that, since

$$h_2(y_1,y_2) = h_2(0,y_2) + \int_0^{y_1} h_{12}(s_1,y_2)ds_1\ , \tag{26}$$

an increase of $h_2(0,y_2)$ shifts the marginal tax rate on the earnings from activity 2 up for all (y_1',y_2') with $y_1' \geq 0$ and $y_2' = y_2$. The total revenue effect is obtained by "summing" all the corresponding conditional revenue terms and this yields $\bar{\mathbf{R}}_2(y_2)$. We refer to $y_2 \mapsto \bar{\mathbf{R}}_2(y_2)$ henceforth as an *unconditional revenue function*. It gives the revenue effect of raising the marginal tax rate on the earnings from activity 2 for all (y_1',y_2') with $y_2' = y_2$. The changes of jointness in the interior of the income space are weighted by the function $y \mapsto \mathbf{R}^J(y)$ with

$$\begin{aligned} \mathbf{R}^J(y) &= \textstyle\int_{y_2}^{\bar{y}_2} \bar{\mathcal{R}}_1(y_1,s_2)ds_2 + \int_{y_1}^{\bar{y}_1} \bar{\mathcal{R}}_2(s_1,y_2)ds_1 \\ &\quad + \textstyle\int_{y_1}^{\bar{y}_1}\int_{y_2}^{\bar{y}_2} \bar{\mathcal{R}}_b(s_1,s_2)ds_2ds_1\ . \end{aligned}$$

The terms $\int_{y_2}^{\bar{y}_2} \bar{\mathcal{R}}_1(y_1,s_2)ds_2$ and $\int_{y_1}^{\bar{y}_1} \bar{\mathcal{R}}_2(s_1,y_2)ds_1$, respectively, reflect that, if at some point y, $h_{12}(y) = h_{21}(y) \neq 0$, then this affects the marginal tax rates on earnings from activity 1, $h_1(y')$, for all y' with $y_1' = y_1$ and $y_2' \geq y_2$ and the marginal tax rates on earnings from activity 2, $h_2(y')$, for all y' with $y_2' = y_2$ and $y_1' \geq y_1$, see Equation (26). When $h_{12}(y) \neq 0$, this also affects the tax burden $h(y')$ for all y' with $y_1' \geq y_1$ and $y_2' \geq y_2$, see Equation (23). This has revenue implications that

are collected in $\int_{y_1}^{\bar{y}_1}\int_{y_2}^{\bar{y}_2}\bar{\mathcal{R}}_b(s_1,s_2)ds_2ds_1$, see Figure 5.

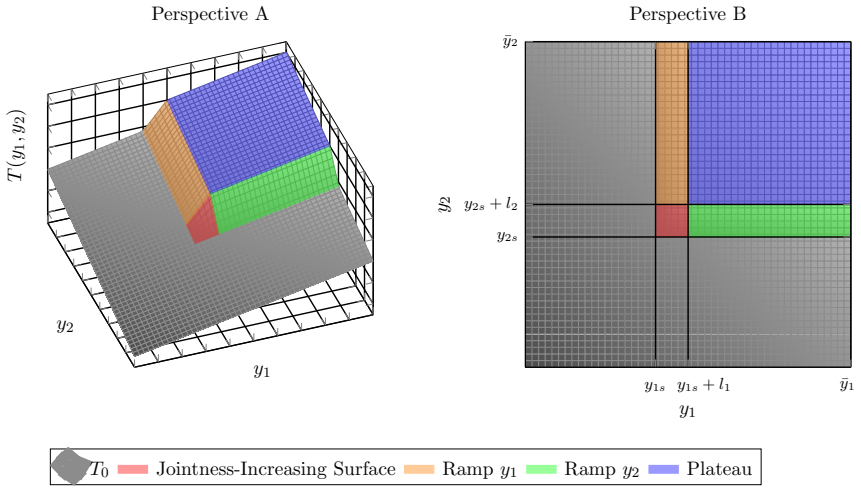
The corresponding expressions in the analysis of welfare implications are the unconditional welfare functions $y_1 \mapsto \bar{\mathbf{W}}_1(y_1)$ with

$$\bar{\mathbf{W}}_1(y_1) := \tfrac{\bar{\mathbf{R}}_1(y_1)}{1-\Delta_\delta^0} - \bar{\mathcal{G}}_1(y_1)f_1(y_1)\ , \quad \bar{\mathcal{G}}_1(y_1) := \textstyle\int_0^{\bar{y}_2}(1-F(y_1 \mid y_2))\mathcal{G}(y_1 \mid y_2)dy_2\ ,$$

and $y_2 \mapsto \mathbf{W}_2(y_2)$, analogously defined. The unconditional welfare functions evaluate the changes of marginal tax rates on the boundaries of the income space. The changes of jointness in the interior of the income space are weighted with

$$\mathbf{W}^J(y) := \tfrac{\mathbf{R}^J(y)}{1-\Delta_\delta^0} - \textstyle\int_{y_1}^{\bar{y}_1}\int_{y_2}^{\bar{y}_2}\bar{g}(s_1,s_2)f(s_1,s_2)ds_2ds_1.$$

Figure 5: A change in jointness



Note: The figure illustrates the implications that a local change of jointness has for marginal tax rates and tax burdens: Jointness changes only in the red rectangle. There, $h_{12}(y) = 1$, and $h_{12}(y) = 0$ elsewhere. In the red area, also marginal tax rates change change since $h_1(y) = y_2 - y_{2s}$ and $h_2(y) = y_1 - y_{1s}$. In the green area, the marginal tax rate on activity 1 does not change, $h_1(y) = 0$, but the marginal tax rate on activity 2 changes since $h_2(y) = \ell_1$. Analogously, in the orange area, there is no change in the marginal tax rate for activity 2, but a change in the marginal rate on activity 1, $h_1(y) = \ell_2$. In the blue area, there is no change in marginal tax rates, but the total tax burden changes since $h(y) = \ell_1\ell_2$.

Implications for local changes of jointness. Suppose that jointness is modified only in some part of the income space: Let $h_{12}^j(y) = \tau_{12}^j$ for $y \in B(y,\ell) := [y_{1s}, y_{1s}+\ell_1] \times [y_{2s}, y_{2s}+\ell_2]$ and $h_{12}^j(y) = 0$, elsewhere. This implies $h^j(y) = 0$ for

$y \notin Y^{top} = \{y \mid y_1 \geq y_{1s}, y_2 \geq y_{2s}\}$. Then

$$(1 - \Delta_\delta^0) R_s(0, h^j) \quad = \quad \tau_{12}^j \int_{B(y,\ell)} \frac{\mathbf{R}^J(y)}{1 - \Delta_\delta^0} \, dy \, .$$

A local change of jointness increases tax revenue if and only if this expression is positive. If a tax system is Pareto efficient, there must not exist self-financing reductions of jointness. This requires that, for all y, $\mathbf{R}^J(y) \geq 0$. Welfare implications are given by

$$W_s(0, h^j) \quad = \quad \tau_{12}^j \int_{B(y,\ell)} \mathbf{W}^J(y) dy. \tag{27}$$

Thus, a local change of jointness is welfare-improving if and only if this expression is positive. In the context of our application, see Section 6, we use this condition to quantify the welfare-implications of reforms that reduce the jointness in the tax treatment of married couples in the US. This analysis shows the empirical relevance of the following scenario: The status quo is locally efficient so that $\mathbf{R}^J(y) > 0$, for all y in a certain range. Also, for some "plausible" specification of welfare weights, $\mathbf{W}^J(y) < 0$. Then, lowering jointness for all y' close to y improves welfare but leads to a loss of tax revenue and hence to a loss of Rawlsian welfare.[37]

5 An impossibility result

In the following we show that it may not be possible to find an optimal tax system in the set of tax systems that satisfy Assumptions 1 and 2. The proof is by contradiction. We first characterize an optimal system and then show that it may not be possible to satisfy all optimality conditions.

Trivially, a necessary condition for the optimality of a tax system is that there is no perturbation that leads to an increase of welfare. A tax system can perturbed in many ways. Hence, there are many such necessary conditions. This raises a question of sufficient conditions. By the following Proposition, we can be sure that a system is optimal, when there is neither an improvement in the class *local changes of one marginal tax rate* nor an improvement in the class *local changes of jointness.*

37. If the status quo has joint and progressive taxation, then a reform that lowers jointness can be interpreted as bringing the tax system closer to individual taxation.

Proposition 5 *The following statements are equivalent*

1. *For all h, $W_s(0,h)=0$.*

2. *For all y, $\mathbf{W}_1(y_1 \mid y_2)=\mathbf{W}_2(y_1 \mid y_2)=\mathbf{W}^J(y)=0$.*

Proposition 5 extends to revenue-maximizing tax systems since tax revenue can be viewed as a degenerate welfare measure, with weights that throughout are equal to 0. Proposition 6 includes a third statement which transcends Proposition 5.

Proposition 6 *The following statements are equivalent*

1. *For all h, $R_s(0,h)=0$.*

2. *For all y, $\mathbf{R}_1(y_1 \mid y_2)=\mathbf{R}_2(y_1 \mid y_2)=\mathbf{R}^J(y)=0$.*

3. *For all y, $\bar{\mathcal{R}}_1(y)=\bar{\mathcal{R}}_2(y)=\bar{\mathcal{R}}_b(y)=0$.*

An inspection of Equation (13) shows that it may not be possible to satisfy all the conditions in statement 3 of Proposition 6. There is no way of having $\bar{\mathcal{R}}_b(y)=0$ in the absence of extensive margin responses.[38] Thus, we cannot take it for granted that a revenue-maximizing tax function exists. To see this, note that $\mathbf{R}^J(y)$ can be written as

$$\begin{aligned}
\mathbf{R}^J(y) &= \int_{y_2}^{\bar{y}_2}\bar{\mathcal{R}}_1(y_1,s_2)ds_2 \;+\; \int_{y_1}^{\bar{y}_1}\Big(\bar{\mathcal{R}}_2(s_1,y_2)ds_1+\int_{y_2}^{\bar{y}_2}\bar{\mathcal{R}}_b(s_1,s_2)ds_2\Big)ds_1 \\
&= \int_{y_2}^{\bar{y}_2}\bar{\mathcal{R}}_1(y_1,s_2)ds_2 \;+\; \int_{y_1}^{\bar{y}_1}\mathbf{R}_2(y_2 \mid s_1)ds_1 \\
&= \int_{y_2}^{\bar{y}_2}\mathbf{R}_1(y_1 \mid s_2)ds_2 \;+\; \int_{y_1}^{\bar{y}_1}\mathbf{R}_2(y_2 \mid s_1)ds_1 \\
&\quad -\int_{y_1}^{\bar{y}_1}\int_{y_2}^{\bar{y}_2}\bar{\mathcal{R}}_b(s_1,s_2)ds_2ds_1\ .
\end{aligned} \tag{28}$$

Now, if $\mathbf{R}_1(y_1 \mid y_2)=\mathbf{R}_2(y_2 \mid y_1)=0$ for all y, so that there is no room for revenue increases with changes of one marginal tax rate, then, by Equation (28), it has to be true that $\mathbf{R}^J(y)<0$, which implies that revenue can be increased by a reduction of jointness. Conversely, if $\mathbf{R}^J(y)=0$, for all y, so that there is no room for revenue increases with changes of jointness, then it must be the case that $\mathbf{R}_i(y_i \mid y_{-i})>0$ for some y and some i, implying that revenue can be increased with a local increase of a marginal tax rate.

38. By contrast, it is possible to have $\bar{\mathcal{R}}_1(y)=\bar{\mathcal{R}}_2(y)=0$ with zero marginal tax rates.

Does the impossibility result extend to welfare-maximizing tax systems? The answer depends on the specification of welfare weights. To see this, suppose that $u(c) = c$ and that there are no extensive margin responses. Therefore, $\bar{\mathcal{R}}_b(y) > 0$, for all y. Still, with welfare weights throughout equal to one and marginal tax rates throughout equal to zero, the following conditions are all satisfied: $\bar{\mathcal{R}}_1(y) = \bar{\mathcal{R}}_2(y) = 0$ and $\mathbf{W}_1(y_1 \mid y_2) = \mathbf{W}_2(y_2 \mid y_1) = \mathbf{W}^J(y) = 0$, for all y. The reason is that the mechanical increases of the tax burden, captured by $\bar{\mathcal{R}}_b(y) = f(y)$, and the welfare losses of those who have to pay higher taxes, equal to $\bar{g}(y)f(y) = f(y)$, exactly cancel each other. With welfare weights that are decreasing in income this is not the case. The insights from the discussion of revenue-maximizing tax systems then extend to welfare-maximizing ones: There is no welfare-maximizing tax function. When marginal tax rates have been set "optimally", the tax system can be improved by lowering jointness. When jointness is "optimal", the tax system can be improved by a change of marginal tax rates.

An interpretation. A reform that involves a change of one marginal tax rate comes with one ramp and one plateau, see Figure 4. By contrast, a reform that involves a change of jointness involves two ramps and one plateau. This implies that combinations of reforms that modify only one marginal tax rate at a time will always have an equal number of ramps and plateaus and can therefore not be used to replicate a modification of jointness. A combination of both reform types can be used to replicate a lump sum change of taxes. To see this, note that Equation (28) can also be written as

$$\int_{y_2}^{\bar{y}_2} \mathbf{R}_1(y_1 \mid s_2)ds_2 + \int_{y_1}^{\bar{y}_1} \mathbf{R}_2(y_2 \mid s_1)ds_1 - \mathbf{R}^J(y) = \int_{y_1}^{\bar{y}_1} \int_{y_2}^{\bar{y}_2} \bar{\mathcal{R}}_b(s_1, s_2)ds_2 ds_1, \quad (29)$$

where the right hand side indicates a lump sum increase of the tax burden in the upper right area of the income space.

Thus, under a tax system that satisfies Assumptions 1 and 2, it is possible to increase taxes at the top in a lump sum fashion by an appropriate combination of continuous reform directions. Welfare-optimality under Assumptions 1 and 2, therefore requires that welfare weights are specified such that collecting these taxes is not desirable. Put differently, when the following two conditions hold (i) addi-

tional lump sum taxes of the rich would improve welfare and (ii) Assumptions 1 and 2 are satisfied, then the status quo cannot be an optimal tax system.

6 Application: The Taxation of Couples

In Section 4 we derived various conditions that can be used to check whether a given tax system can be improved in terms of efficiency and social welfare. We now use these conditions to evaluate reform options for the current tax treatment of married couples in the US. The overarching question that motivates our analysis is whether it is desirable to "move towards individual taxation". We analyze two ways of moving closer to individual taxation, reforms that lead to lower marginal tax rates on secondary earnings and reforms that reduces the tax system's degree of jointness. We clarify the implications that such reforms would have for alternative measures of social welfare. Specifically, we consider a Rawlsian social welfare function which assigns weights only to the poorest individuals who have no market income at all. We also use utilitarian welfare with equals weights, which – in combination with the assumption that utility increases linearly in consumption – indicates how the economy's overall surplus, defined as total output minus effort costs, is affected by the reform. A "Near-Rawlsian" welfare function concentrates weights at the bottom of the income distribution, without being lexicographic. Finally, we explore a "Feminist" social welfare function which has weights that are increasing in the secondary earner's income share.[39] Broadly, our analysis shows that, lowering tax rates on secondary earnings away from the bottom of the income distribution can be Pareto-improving, hence approved by all these welfare measures. Realizing these Pareto-improvements requires a lot of fine-tuning however: the tax cuts for secondary earnings apply only if both primary and secondary earnings lie in appropriate ranges. Lowering marginal tax rates at the bottom is approved by a near-Rawlsian welfare measure but rejected by a Rawlsian one. We find, moreover, that reductions of jointness provoke a conflict between "the poor" and "working middle-class women", they are approved by "Feminist" welfare measures and rejected by Rawlsian ones. Finally, we characterize optimal tax rates at the top and do so separately for primary and secondary earnings. The main result is that

39. For details on the formal specification of welfare weights, see Table L5.

"rich women" should be taxed at lower rates than "rich men", even if behavioral responses of the two are the same.

In this analysis we attach a specific interpretation to the framework that was developed in the previous sections. What previously was an abstract "agent" is now a married couple consisting of a spouse with index 1 and spouse with index 2 who, respectively have earnings of y_1 and y_2.[40] We assume that there is a status quo tax function T with progressive and joint taxation. For simplicity, we let $u(c) = c$.[41] We impose the assumption that the status quo does not give rise to mass points.[42] We finally impose the assumption that effort cost functions take the iso-elastic form described in Equation (2). Under these assumptions the function $y \mapsto \bar{\mathcal{R}}_b(y)$ is given by (16) and the functions $y \mapsto \bar{\mathcal{R}}_1(y)$, $y \mapsto \bar{\mathcal{R}}_2(y)$ are given by (18). In Appendix L.2 we discuss in detail how we calibrate the different types of revenue functions under the assumption of a piece-wise linear tax system. The calibration requires three types of inputs: Information on the distribution of primary and secondary earnings, information on the status quo tax system, and estimates of the elasticities that capture the behavioral responses to taxation.

Data on primary and secondary earnings. We draw on detailed household micro data from the Annual Social and Economic Supplement of the CPS (CPS-ASEC).[43] The detailed demographic characteristics in the CPS allow us to obtain information on the (conditional) income distributions of primary and secondary earners in married couples.

The tax system. Information on tax rates is obtained through the NBER TAXSIM (v32) microsimulation model.[44] By combining CPS data with TAXSIM, we esti-

40. We assign the index 2 to the spouse who has lower earnings.

41. Like much of the literature, we treat a couple's disposable income as a public good. An extensive discussion of alternative assumptions can be found in Bierbrauer et al. (2023).

42. We use a data set where people mostly have labor incomes and thus plausibly have only very limited flexibility to adjust their earnings so as to bunch at the kink points. This would be different with a sample of self-employed people.

43. See Flood et al. (2021) and https://cps.ipums.org for a detailed description of CPS data. Appendix L.1 provides details on the data preparation. We use CPS data because it provides separate demographic and earnings information for both spouses. In contrast, the tax return micro data (SOI-PUF) from the Internal Revenue Service (IRS) used in Bierbrauer, Boyer, and Peichl (2021) does not contain this information (except for the year 1974; see Figure L3 for a comparison).

44. See Feenberg and Coutts (1993) and https://users.nber.org/~taxsim/ for detailed information on the TAXSIM microsimulation model, and Appendix L.1 for details on its combination

mate tax rates and tax liabilities under the status quo at the household level. This allows us to capture complexities in the tax code that go beyond the information embedded in statutory marginal tax rates. We focus on the tax system as of 2019.

Behavioral responses. In a baseline scenario, we assume intensive-margin elasticities of $\epsilon_1 = 0.25$ for primary and $\epsilon_2 = 0.75$ for secondary earners, following central estimates from the empirical public finance literature.[45] Alternative parameterizations are reported in Table L2. For the extensive margin, we assume $L_E(y) = 0.2$, for all y, in the baseline scenario. For robustness, we also provide results without extensive margin responses, i.e., a value of $L_E = 0$.[46]

6.1 Lowering marginal tax rates on secondary earnings.

Our analysis of reforms towards individual taxation begins by examining reforms that reduce marginal tax rates for secondary earnings. In particular, we test whether such reforms are self-financing depending on where exactly the tax cuts take place.

Pareto-improvements. The conditional revenue function $y \mapsto \mathbf{R}_2(y_2 \mid y_1)$ gives the revenue implications from local changes of the marginal tax rate on secondary earnings in the vicinity of y_2, while holding primary earnings fixed at y_1. When $\mathbf{R}_2(y_2 \mid y_1) < 1$ this indicates that lowering marginal tax rates on secondary earnings would be self-financing and hence Pareto-improving at a particular combination of primary earnings y_1 and secondary earnings y_2. To see whether such combinations can be found empirically, Figure 6 contains plots of

$$\mathbf{R}_2(y_2 \mid B_1) \quad = \quad \int_{y_1 \in B_1} \mathbf{R}_2(y_2 \mid y_1) f_1^e(y_1) dy_1$$

for alternative specifications of the conditioning primary earnings bracket B_1.[47] For Figure 6, the different values of B_1 correspond to the deciles of the marginal

with CPS data.

45. See, e.g., Blundell and Macurdy (1999), Blau and Kahn (2007), Eissa and Hoynes (2004), LaLumia (2008), Bargain, Orsini, and Peichl (2014).

46. As shown in part K of the Supplement, L_E is a sum of various extensive margin elasticities that, moreover, are a composite of fiscal externalities and behavioral responses each. Thus, we have less guidance from the empirical literature for this parameter.

47. We condition on income brackets instead of exact primary earnings, since we require enough secondary earnings data points to construct conditional revenue functions. For additional details on the estimation, see Appendix L.2. Figures L6 – L25 report all relevant ingredients.

distribution of primary earnings.[48] For comparison the Figure also includes plots of $y_1 \mapsto \mathbf{R}_1(y_1 \mid B_2)$, for alternative specifications of B_2 that correspond to the deciles of the secondary earnings distribution.

Figure 6a shows that, conditional on $y_1 \in B_1$, marginal tax rates on secondary earnings are inefficiently high in the upper deciles of the secondary earnings distribution. This is most pronounced for "middle class" primary earnings. For instance, consider primary earnings above the median (Q6). The associated conditional revenue function is positive and monotonically decreasing until secondary earnings exceed \$70,000. Then, the function turns negative indicating inefficiently high marginal tax rates on secondary earnings. Hence, a reduction of secondary earners marginal tax rates for all couples with secondary earnings above \$70,000 and primary earnings in the sixth income decile would be self-financing.

A comparison with the unconditional revenue function

$$\bar{\mathbf{R}}_2(y_2) \quad = \quad \int_0^{\bar{y}_1} \mathbf{R}_2(y_2 \mid y_1) f_1(y_1) dy_1$$

shows that the conditioning is important. Inefficiencies can only be eliminated with tax cuts that are tailored to the relevant decile of the primary earnings distribution. To see this, note that the unconditional revenue function at secondary earnings of \$70,000 is positive. Thus, cutting marginal tax rates on secondary earnings for all couples with y_2 close to \$70,000 would not be self-financing.

When applied to primary earnings, this test does not identify inefficiencies (see Figure 6b). Thus, it is not the case that marginal tax rates are inefficiently high across the board and by implication also for secondary earnings. They are too high only for secondary earnings which are close to primary earnings.

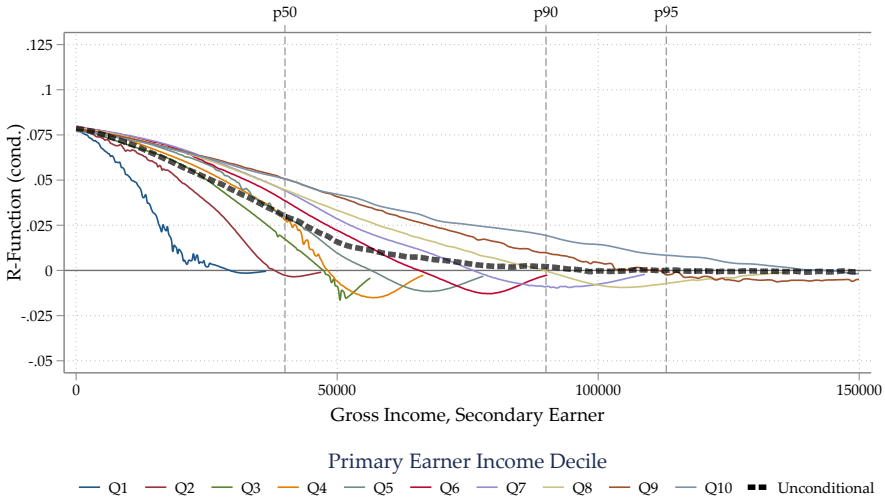
In the Appendix, we replicate the analysis by conditioning on earnings quintiles (Figure L29), vingtiles (Figure L30), and statutory income tax brackets (Figure L31). In line with the main figure, the magnitude of inefficiencies for secondary earners is largest in the middle of the primary earnings income distribution or, alternatively, for couples with primary earners in "middle" tax brackets.

Lowering marginal tax rates at the bottom. For income ranges in which we do not find inefficiencies, it might still be the case that reducing marginal tax rates

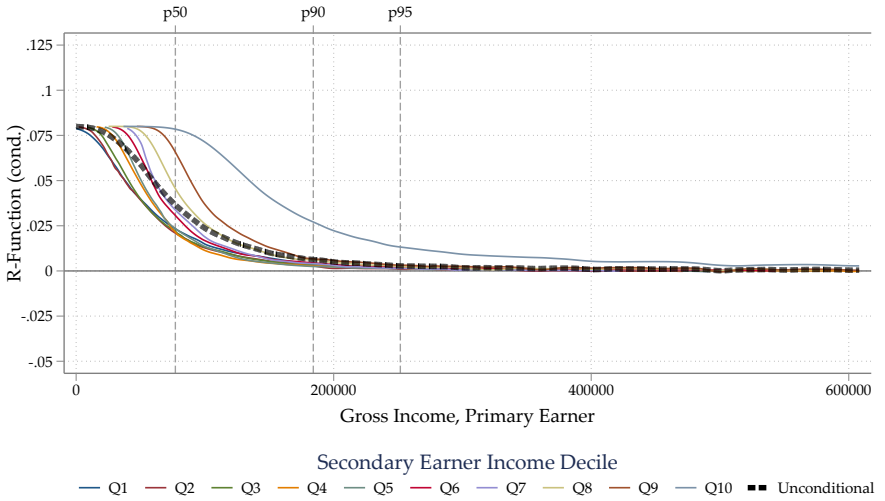
48. See Table L3 for the corresponding income thresholds.

Figure 6: Conditional revenue functions, deciles (2019)

(a) Secondary earners



(b) Primary earners



Note: This figure shows conditional revenue functions for secondary (resp. primary) earners in married dual earner couples conditional on primary (resp. secondary) earnings income deciles in Panel a (resp. b) as of 2019. All estimates are based on tax units with non-negative gross income in which both spouses are between 25 and 55 years old. Our pre-tax gross income variable of interest contains wage income, farm income, business income, income from dividends, income from interest, income from rent, and retirement income. We take behavioral responses at the intensive and extensive margin into account. We assume intensive margin elasticities of 0.25 (resp. 0.75) for primary (resp. secondary) earners and an extensive margin elasticity of 0.2. The figure also displays modified unconditional revenue functions for secondary and primary earners where unconditional revenue functions have been scaled by 0.1 to facilitate comparability with the conditional revenue functions. Results without extensive margin responses are shown in Figure L26. Results for alternative elasticities (Table L2) are shown in Figures L27 and L28.
Source: Own calculations based on CPS-ASEC (2019).

on secondary earnings is welfare-improving under some measures of social welfare. In a situation where

$$\mathbf{W}_2(y_2 \mid y_1) \quad < \quad 0 \quad < \quad \mathbf{R}_2(y_2 \mid y_1),$$

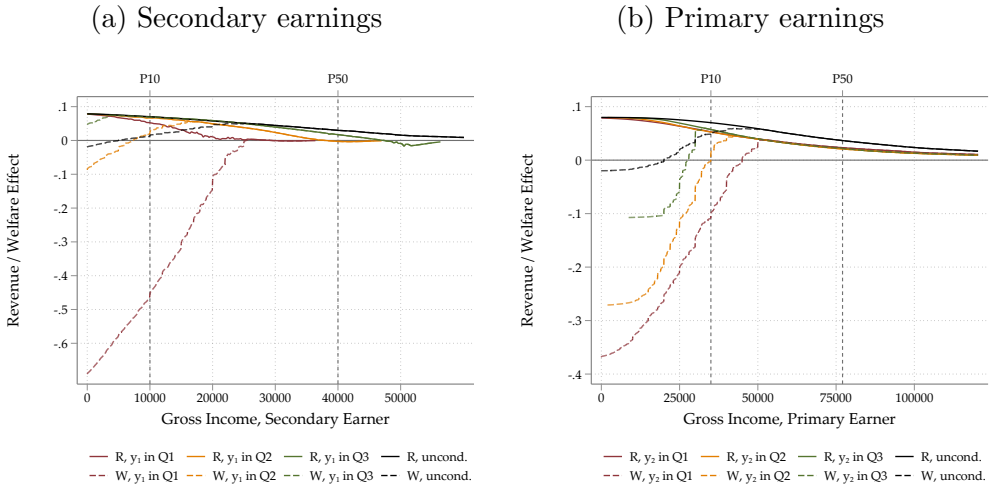
lowering marginal tax rates on secondary earnings is costly in terms of revenue, hence rejected by Rawlsian welfare, but desirable from the perspective of a non-Rawlsian welfare measure shaping the conditional welfare function $y \mapsto \mathbf{W}_2(y_2 \mid y_1)$.

Figure 7 shows revenue and welfare functions conditional on primary earnings below the 30th percentile of the primary earnings distribution. It also presents the conditional functions that indicate whether or not lowering marginal tax rates on primary earnings is desirable. The solid lines are the conditional revenue functions which describe the revenue effects of the tax change, or equivalently, the welfare effects under a Rawlsian welfare function. The dashed lines show the conditional welfare functions for near-Rawlsian welfare weights. The figure also displays the revenue and welfare effects of reforms that lower marginal tax rates on secondary or primary earnings unconditionally (black lines).

The figure shows that lowering marginal tax rates at the bottom is costly in terms of revenue. The assumption that losses of revenue lead to a downward adjustment of the basic transfer to individuals with no income implies that the reform is akin to an introduction or strengthening of workfare elements: less unconditional transfers and more disposable income in return for work effort. The loss of revenue implies that such a reform is rejected by a Rawlsian welfare function. By contrast, near-Rawlsian welfare goes up when marginal tax rates are lowered, and this applies both to primary and to secondary earnings. Again, the conditioning on spousal earnings is important for this: the unconditional near-Rawlsian welfare function turns negative only over a narrow range at the very bottom.[49]

49. When excluding extensive margin responses, the unconditional near-Rawlsian welfare function remains positive throughout, whereas the conditional counterparts take negative values (Figure L32), driven by differences in average (un)conditional welfare weights (Figure L33).

Figure 7: Welfare effects of reforms at the bottom (2019)



Note: This figure shows the revenue and welfare effects for secondary (resp. primary) earners in married dual earner couples conditional on primary (resp. secondary) earnings income deciles at the bottom of the income distribution in Panel a (resp. b) as of 2019. All estimates are based on tax units with non-negative gross income in which both spouses are between 25 and 55 years old. Our pre-tax gross income variable of interest contains wage income, farm income, business income, income from dividends, income from interest, income from rent, and retirement income. We take behavioral responses at the intensive and extensive margin into account. We assume intensive margin elasticities of 0.25 (resp. 0.75) for primary (resp. secondary) earners and an extensive margin elasticity of 0.2. The figure displays conditional revenue functions (solid lines, also displayed in detail in Figure 6) and the welfare effect based on near-Rawlsian welfare weights according to Table L5 (dashed lines). Black lines indicate unconditional revenue and welfare functions. Results without extensive margin responses are shown in Figure L32. Results for different elasticity scenarios in line with Table L2 are shown in Figure L34.
Source: Own calculations based on CPS-ASEC (2019).

6.2 Reducing jointness

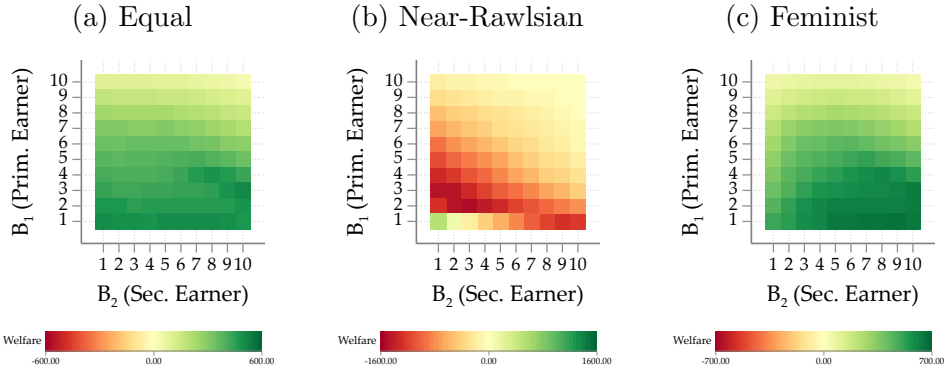
An alternative way to think about reforms that move towards individual taxation is to consider reductions in jointness. We consider changes of jointness tailored to specific combinations of income brackets for primary and secondary earnings and evaluate them for equal (surplus-maximizing) weights, near-Rawlsian welfare weights, and "Feminist" welfare weights. Our evaluation uses the functional derivative $W_s(0, h^j)$ characterized in Equation (27). If $W_s(0, h^j) > 0$ (resp. < 0), then it is desirable to increase (resp. decrease) jointness, that is, the tax function's cross-derivative should be raised (resp. lowered) for all earnings within the box with lower-left corner $y(y_1, y_2)$ and upper right corner $(y_1 + \ell_1, y_2 + \ell_2)$.

In our empirical implementation, we let $\ell_1 = \ell_2 = 500$ and compute $\tau_{12}^j \int_{B(y,\ell)} \mathbf{W}^J(y)$ for $\tau_{12}^j = -0.01$, corresponding to a one percentage point reduction of the tax function's cross derivative. We select the thresholds y_1 for primary earnings based on the deciles of the primary earner income distribution, and the thresholds for secondary earnings y_2 based on the deciles of the distribution of secondary earnings

conditional on the respective primary earner income decile.[50]

Figure 8 shows the results. Total surplus unambiguously goes up if jointness is reduced. The same is true for "Feminist" welfare, but with more pronounced welfare gains in the range of high secondary earnings. The near-Rawlsian welfare measure rejects reductions in jointness, except at the very bottom. This is consistent with the previous finding that, at the bottom, the near-Rawlsian welfare measure approves reductions of marginal tax rates both for primary and for secondary earnings.[51]

Figure 8: Welfare implications of decreasing jointness (2019)



Note: This figure displays the welfare implications of decreasing jointness as of 2019 by decreasing secondary and primary marginal tax rates conditional on particular brackets of primary earnings and secondary earnings. Bracket thresholds are shown in Table L6. The reform applies to a bracket of length $l(B_1) = l(B_2) = 500$ and has the magnitude $\tau = -0.01$. The figure distinguishes between three different forms of welfare weights (see Table L5) and is based on intensive margin responses according to the baseline elasticity scenario (see Table L2). All results are shown including extensive margin responses. Results for different elasticity scenarios and without extensive margin responses are shown in Figures L38 and L39.
Source: Own calculations based on CPS-ASEC (2019).

6.3 Optimal top tax rates

Under the premises of Proposition 4 top tax rates for primary and secondary earners are given by

$$\frac{t_i^*}{1-t_i^*} = (1-\bar{g}_i)\,\frac{1}{\bar{\alpha}_i}\frac{1}{\varepsilon_i} \qquad i = 1, 2.$$

We seek to trace out the implications of differences in the Pareto-coefficients for primary and secondary earnings for top tax rates. We therefore assume, for ease of exposition, that "rich people" have welfare weights of zero, $g_1 = g_2 = 0$. As in the

50. The bracket thresholds are shown in Table L6.
51. In Appendix L.4 we provide insights on how the aggregate welfare effect shown in Figure 8 can be decomposed into mechanical effects, behavioral responses and welfare weight contributions.

previous sections, in our baseline, we assume that $\varepsilon_1 = 0.25$ and $\varepsilon_2 = 0.75$.[52]

Pareto coefficients. We estimate Pareto coefficients for primary and secondary earnings among jointly filing tax units using tabulated income tax data from the Statistics of Income (SOI) program of the Internal Revenue Service (IRS).[53] We use the tabulated tax return data and apply generalized Pareto interpolation techniques by Blanchet, Fournier, and Piketty (2022) to estimate properties of the wage income distribution. In the main text, we focus on 2018, the most recent available year for this data. Panel A of Table 1 shows the estimated Pareto coefficients at the 95th and 99th percentile of the wage income distribution. At both percentiles, the Pareto coefficient of the primary earner wage income distribution is around 40 percent lower than the respective coefficient for secondary earners.

Table 1: Pareto coefficients and optimal tax rates (2018)

	P95		P99	
	Primary	Secondary	Primary	Secondary
Panel A: Pareto coefficients				
Wages on W2 form	1.83	3.06	1.75	3.12
Panel B: Optimal tax rates				
Elasticity = .25	69%	57%	70%	56%
Elasticity = .5	52%	40%	53%	39%
Elasticity = .75	42%	30%	43%	30%

Note: This table shows Pareto coefficients and optimal top tax rates. Panel A displays Pareto coefficients for primary and secondary earners based on a generalized Pareto interpolation using tabulated data on wages on W2 forms for joint return taxpayers with wage income. We distinguish between interpolations at the 95th and the 99th percentile. Panel B displays optimal top tax rates associated with these Pareto coefficients. We distinguish between different elasticities.
Source: Own calculations based on SOI Tax Stats - Individual Information Return Form W-2 (2018)

Results. Based on the estimated Pareto coefficients, Panel B of Table 1 displays optimal top tax rates for different elasticities. Even under the assumption that $\varepsilon_1 = \varepsilon_2$ at the top of the income distribution, the optimal tax rates for primary earners are around 12 percentage points higher than the one for secondary earners.[54]

52. We proceed under Assumption 3. In particular, we take elasticities to be constant over the relevant part of the income distribution. An implication is that the Pareto coefficient does not vary with the tax rates at the top.

53. For details on the data source and preparation, see Appendix L.1.2.

54. Supplementary results on the evolution of optimal top tax rates over time, and heterogeneity between single earner couples and dual earner couples are relegated to Appendix L.5.

7 Concluding remarks

Problems of one-dimensional screening are well understood. In particular, when preferences are linear in disposable income they can be articulated as problems of virtual surplus- or welfare-maximization. This approach takes care of resource and incentive constraints and encapsulates them in a modified policy objective. Outcomes can then be evaluated as if this modified objective was the true objective and as if there were no resource or incentive constraints to begin with. The derivation of this virtual policy objective involves an integration by parts to capture how local distortions of an allocation affect payoffs or information rents further up in the type distribution.[55]

This paper's analysis is inspired by this approach. To be clear, the problem here is one of multi-dimensional screening, preferences are not a priori restricted to be quasi-linear and we do not present a virtual policy objective in closed form. Instead, we derive functional derivatives that capture the welfare implications of tax reforms. Getting there involves integration by parts to captures how local changes in marginal tax rates or jointness affect tax burdens and behavioral responses further up in the income distribution. These functional derivatives can be interpreted as the derivatives of a virtual policy objective: They indicate whether a reform brings the actual measure of welfare up or down. The logic of this constructions transcends the specific setup on which this paper's analysis is based. We therefore conjecture that it has the same portability to other problems of multi-dimensional screening as the closed form virtual surplus measure has for one-dimensional screening problems.

55. The notion of virtual surplus has been introduced by Myerson (1981) in the context of an auction design problem and then been used more widely in other applications. Rochet (2024) points out that an analogue is missing for problems of multi-dimensional screening.

References

Ahlvik, L., M. Liski, and M. Mäkimattila. 2024. "Pigouvian Income Taxation." *CESifo Working Paper No. 11174.*

Akbarpour, M., P. Doligalski, P. Dworczak, and S. Kominers. 2025. "Optimal Redistribution via Income Taxation and Market Design." *Working Paper.*

Ales, L., and C. Sleet. 2022. "Optimal Taxation of Income-Generating Choice." *Econometrica* 90 (5): 2397–2436.

Alves, C.B., C.E. da Costa, F. Lobel, and H. Moreira. 2024. "Intrahousehold inequality and the joint taxation of household earnings." *Journal of Public Economics* 239:105208.

Armour, P., R. V. Burkhauser, and J. Larrimore. 2016. "Using the Pareto distribution to improve estimates of topcoded earnings." *Economic Inquiry* 54 (2): 1263–1273.

Bargain, O., M. Dolls, H. Immervoll, D. Neumann, A. Peichl, N. Pestel, and S. Siegloch. 2015. "Tax policy and income inequality in the United States, 1979–2007." *Economic Inquiry* 53 (2): 1061–1085.

Bargain, O., K. Orsini, and A. Peichl. 2014. "Comparing labor supply elasticities in europe and the united states new results." *Journal of Human Resources* 49 (3): 723–838.

Bergstrom, K., and W. Dodds. 2021. "Optimal taxation with multiple dimensions of heterogeneity." *Journal of Public Economics* 200:104442.

———. 2025. "A General Theory of Inverse Welfare Functions." *Mimeo.*

Bick, A., and N. Fuchs-Schündeln. 2017. "Taxation and Labour Supply of Married Couples across Countries: A Macroeconomic Analysis." *Review of Economic Studies* 85 (3): 1543–1576.

Bierbrauer, F.J., P.C. Boyer, and E. Hansen. 2023. "Pareto-improving tax reforms and the Earned Income Tax Credit." *Econometrica* 91 (3): 1077–1103.

Bierbrauer, F.J., P.C. Boyer, and A. Peichl. 2021. "Politically feasible reforms of nonlinear tax systems." *American Economic Review* 111 (1): 153–91.

Bierbrauer, F.J., P.C. Boyer, A. Peichl, and D. Weishaar. 2023. *The taxation of couples.* CEPR Discussion Paper 18138.

———. 2024. "The political economy of joint taxation." *CESifo Working Paper Series 11728.*

Blanchet, T., J. Fournier, and T. Piketty. 2022. "Generalized Pareto Curves: Theory and Applications." *Review of Income and Wealth* 68 (1): 263–288.

Blau, F. D., and L. M. Kahn. 2007. "Changes in the labor supply behavior of married women: 1980–2000." *Journal of Labor Economics* 25 (3): 393–438.

Blundell, R., and T. Macurdy. 1999. "Chapter 27 - Labor Supply: A Review of Alternative Approaches," edited by Orley C. Ashenfelter and David Card, 3:1559–1695. Handbook of Labor Economics. Elsevier.

Boerma, J., A. Tsyvinski, and A.P. Zimin. 2022. *Bunching and Taxing Multidimensional Skills.* NBER Working Paper 30015.

Borella, M., M. De Nardi, and F. Yang. 2023. "Are Marriage-Related Taxes and Social Security Benefits Holding Back Female Labour Supply?" *Review of Economic Studies* 90 (1): 102–131.

Boskin, M. J., and E. Sheshinski. 1983. "Optimal tax treatment of the family: Married couples." *Journal of Public Economics* 20 (3): 281–297.

Brett, C. 2009. "Optimal nonlinear taxes for families." *International Tax and Public Finance* 14 (3): 225–261.

Cremer, H., J.-M. Lozachmeur, and P. Pestieau. 2012. "Income taxation of couples and the tax unit choice." *Journal of Population Economics* 25 (2): 763–778.

Diamond, P.A. 1998. "Optimal Income Taxation: An Example with a U-Shaped Pattern of Optimal Marginal Tax Rates." *American Economic Review* 88:83–95.

Eissa, N., and H. W. Hoynes. 2004. "Taxes and the labor market participation of married couples: the Earned Income Tax Credit." *Journal of Public Economics* 88 (9-10): 1931–1958.

Eissa, N., H.J. Kleven, and C.T. Kreiner. 2008. "Evaluation of four tax reforms in the United States: Labor supply and welfare effects for single mothers." *Journal of Public Economics* 92 (3): 795–816.

Feenberg, D. R., and E. Coutts. 1993. "An Introduction to the TAXSIM Model." *Journal of Policy Analysis and Management* 12 (1): 189–194.

Ferey, A., B. B. Lockwood, and D. Taubinsky. 2024. "Sufficient Statistics for Nonlinear Tax Systems with General Across-Income Heterogeneity." *American Economic Review* 114 (10): 3206–49.

Flood, S., M. King, R. Rodgers, S. Ruggles, J. R. Warren, and M. Westberry. 2021. "Integrated Public Use Microdata Series, Current Population Survey: Version 9.0 [dataset]."

Gayle, G.-L., and A. Shephard. 2019. "Optimal taxation, marriage, home production, and family labor supply." *Econometrica* 87 (1): 291–326.

Golosov, M., and I. Krasikov. 2025. "The Optimal Taxation of Couples." *Quarterly Journal of Economics* 140 (3): 2163–2211.

Guesnerie, R. 1995. *A Contribution to the Pure Theory of Taxation.* Cambridge University Press.

Guner, N., R. Kaygusuz, and G. Ventura. 2012. "Taxation and household labour supply." *Review of Economic Studies* 79 (3): 1113–1149.

Hammond, P. 1979. "Straightforward Individual Incentive Compatibility in Large Economies." *Review of Economic Studies* 46:263–282.

Holter, H., D. Krüger, and S. Stepancuk. 2023. *Until the IRS Do Us Part: (Optimal) Taxation of Households.* Mimeo. University of Pennsylvania.

Immervoll, H., H. J. Kleven, C. T. Kreiner, and N. Verdelin. 2011. "Optimal tax and transfer programs for couples with extensive labor supply responses." *Journal of Public Economics* 95 (11): 1485–1500.

Jacquet, L., and E. Lehmann. 2021a. *How to Tax Different Incomes?* CEPR Discussion Paper 16571.

———. 2021b. "Optimal income taxation with composition effects." *Journal of the European Economic Association* 19 (2): 1299–1341.

Kleven, H. J., C. T. Kreiner, and E. Saez. 2009. "The Optimal Income Taxation of Couples." *Econometrica* 77 (2): 537–560.

LaLumia, S. 2008. "The effects of joint taxation of married couples on labor supply and non-wage income." *Journal of Public Economics* 92 (7): 1698–1719.

Larrimore, J., R. V. Burkhauser, S. Feng, and L. Zayatz. 2008. "Consistent cell means for topcoded incomes in the public use March CPS (1976–2007)." *Journal of Economic and Social Measurement* 33 (2-3): 89–128.

Malkov, E. 2021. *Optimal Income Taxation of Singles and Couples.* Unpublished Manuscript.

Milgrom, P., and I. Segal. 2002. "Envelope Theorems for Arbitrary Choice Sets." *Econometrica* 70 (2): 583–601.

Milgrom, P., and C. Shannon. 1994. "Monotone Comparative Statics." *Econometrica* 62 (1): 157–180.

Myerson, R. B. 1981. "Optimal auction design." *Mathematics of Operations Research* 6 (1): 58–73.

Neisser, C. 2021. "The elasticity of taxable income: A meta-regression analysis." *Economic Journal* 131 (640): 3365–3391.

Piketty, T. 2003. "Income Inequality in France, 1901-1998." *Journal of Political Economy* 111 (5): 1004–1042.

Piketty, T., and E. Saez. 2003. "Income Inequality in the United States, 1913-1998." *Quarterly Journal of Economics* 118 (1): 1–41.

Rochet, J.-C. 2024. "Multidimensional screening after 37 years." *Journal of Mathematical Economics* 113:103010.

Saez, E., J. Slemrod, and S.H. Giertz. 2012. "The elasticity of taxable income with respect to marginal tax rates: A critical review." *Journal of Economic Literature* 50 (1): 3–50.

Schroyen, F. 2003. "Redistributive taxation and the household: the case of individual filings." *Journal of Public Economics* 87 (11): 2527–2547.

Spiritus, K., E. Lehmann, S. Renes, and F. Zoutman. 2025. "Optimal taxation with multiple incomes and types." *Theoretical Economics* 20 (2): 543–581.

Appendix

A Proof of Proposition 1

To prove Proposition 1 we verify the following statements:

(i) For all θ, the choice problem in (1) has a solution. The value function $\theta \mapsto V(\theta)$ is continuous. The allocation $\theta \mapsto y^*(\theta)$ can be implemented with a tax function that is continuous.

(ii) For every ω, there is a partition Φ^{12}, Φ^1, Φ^2, Φ^0 of $\mathbb{R}^2_0$ s.t. φ in the interior of Φ^{12} implies $U^{12}(\theta) > \max\{U^1(\theta), U^2(\theta), U^0\}$; analogously for Φ^1, Φ^2 and Φ^0.

(iii) For every $y_{-i} \in \mathbb{R}^+_0$, there is a partition $(\Omega_{i,k})_{k\in K}$, $K \subset \{1, \ldots, m\}$, of Ω_i so that ω_i in the interior of $\Omega_{i,k}$ implies that $v_k(\omega_i \mid y_{-i}) > v_l(\omega_i \mid y_{-i})$ for any $l \neq k$. If $\Omega_{i,k}$ and $\Omega_{i,l}$ are non-empty, there is $\hat{\omega}_i^{kl}$ so that $v_k(\omega_i \mid y_{-i}) > v_l(\omega_i \mid y_{-i})$, for $\omega_i < \hat{\omega}_i^{kl}$ and $v_k(\omega_i \mid y_{-i}) < v_l(\omega_i \mid y_{-i})$, for $\omega_i > \hat{\omega}_i^{kl}$.

(iv) For almost all θ, the choice problem in (1) has a unique solution.

(v) For almost all θ and all $i \in \{1, 2\}$, if $y_i^*(\theta) > 0$, then either there is strict bunching

$$\begin{aligned} & u'\Big(C(y^*(\theta))\Big)\Big(1 - \lim_{\epsilon\to 0^+} T_{y_i}(y_i^*(\theta) + \epsilon, y_{-i}^*(\theta))\Big) \\ < \quad & k_{y_i}(y_i^*(\theta), \omega_i) \\ < \quad & u'\Big(C(y^*(\theta))\Big)\Big(1 - \lim_{\epsilon\to 0^-} T_{y_i}(y_i^*(\theta) + \epsilon, y_{-i}^*(\theta))\Big)\,, \end{aligned} \tag{30}$$

or a first order condition is satisfied,

$$u'\Big(C(y^*(\theta))\Big)\Big(1 - T_{y_i}(y^*(\theta))\Big) \;=\; k_{y_i}(y_i^*(\theta), \omega_i)\,. \tag{31}$$

A.1 Parts (i) and (ii)

A.1.1 Existence and Part (ii)

We seek to show that, for all types θ, the utility-maximization problem has a solution. We decompose this overall problem into a set of subproblems and show

that they have a solution. As a final step, we show that the solution to the overall problem can be defined with reference to the subproblems.

Optimal choices conditional on paying both fixed costs. The best choice conditional on paying the fixed costs φ_1 and φ_2 solves

$$\max_{y_1,y_2\in\mathcal{Y}} \; u(C(y_1,y_2)) - k_1(y_1,\omega_1) - k_2(y_2,\omega_2) \; .$$

We first consider an auxiliary problem: By Assumption 1, there is a finite partition $\mathcal{P} = \{P_1,\ldots,P_m\}$ of $\mathcal{Y}$ so that, for all $j \in \{1,\ldots,m\}$, T is continuously differentiable over the interior of P_j. For any set P in $\mathcal{P}$, denote its closure by $\bar{P}$. Denote the collection of all such sets by $\bar{\mathcal{P}}$. Define the tax function on the boundary of any $\bar{P}$ by taking the appropriate limit from the interior of $\bar{P}$. For any such $\bar{P}$, there is now a well defined problem: Choose $y \in \bar{P}$ to maximize utility. This is the maximization of a continuous function over a compact set, hence a problem with a solution. We write $Y^{12}_{\bar{P}}(\omega)$ for the set of maximizers and $v^{12}_{\bar{P}}(\omega)$ for the value function. We now define the value function of the auxiliary problem as

$$v^{12}_{\bar{\mathcal{P}}}(\omega) \quad = \quad \max_{\bar{P}\in\bar{\mathcal{P}}} \quad v^{12}_{\bar{P}}(\omega)$$

and the corresponding set of maximizers by $Y^{12}_{\bar{\mathcal{P}}}(\omega)$.

We now argue that the solution of the auxiliary problem coincides with the solution to the problem of interest. By including points of closure to the auxiliary optimization problem, we have made some choices available that are not in the choice set of the problem of interest. These additional points of closure can be removed, however, without affecting the solution to the auxiliary problem. To see this, consider a closure point y that belongs to some $\bar{P}$ but not to P. Also note that, since $\mathcal{P}$ is a partition, y needs to belong to some element P' of $\mathcal{P}$. Suppose first that y is a kink, i.e. a point at which the tax function is continuous, but not differentiable. Then the utility that is realized when choosing y from $\bar{P}$ is equal to the utility that is realized when choosing y from P'. Therefore removing y from $\bar{P}$ does not lower the attainable level of utility. Suppose now that y is a notch, i.e. a point at which the tax function is discontinuous, but not differentiable. Then, by Assumption 2, the utility that is realized when choosing y from $\bar{P}$ is lower than the

utility that is realized when choosing y from P'. Again, removing y from $\bar{P}$ does not lower the level of utility that is attainable in the problem of interest.

We denote by $Y^{12}(\omega)$ the set of maximizers for the above subproblem and by $v^{12}(\omega)$ the corresponding value function.

Optimal choices conditional on paying only the fixed cost for one activity. The best choice conditional on paying only the fixed costs φ_1 solves

$$\max_{y_1 \in \mathcal{Y}_1} u(C(y_1, 0)) - k(y_1, \omega_1) \ .$$

one can show along the same lines as in the preceding paragraph that this problem has a solution. Let $Y^1(\omega_1)$ be the (not necessarily single-valued) argmax of this optimization problem. Denote the corresponding value function by $v^1(\omega_1)$. We use analogous notation for the optimal choice conditional on paying only the fixed costs φ_2.

Extensive margin choices/ Part (ii). Paying both fixed costs is optimal when the following three inequalities are all satisfied: $v^{12}(\omega) - v^0 \geq \varphi_1 + \varphi_2$, where $v^0 = u(b)$, $v^{12}(\omega) - v^1(\omega_1) \geq \varphi_2$, and $v^{12}(\omega) - v^2(\omega_2) \geq \varphi_1$. For later reference, let $\Phi^{12}(\omega)$ be the set of all φ satisfying these three inequalities. We define $\Phi^1(\omega)$ and $\Phi^2(\omega)$ analogously. If the distribution of φ conditional on ω is atomless, then there is only zero mass of types who are indifferent between these choices.

The sets $\Phi^{12}(\omega)$, $\Phi^1(\omega)$ and $\Phi^2(\omega)$ can be drawn in φ_2-φ_1-space (see Figure 3). Their boundaries are linear because the inequalities defining them are linear in φ_1 and φ_2.

A.1.2 Continuity of the value function

We show continuity in ω_i, for fixed values of $\varphi = (\varphi_1, \varphi_2)$ and ω_{-i}. The argument below can be easily adapted to also establish continuity in these other dimensions. Fix φ and ω_{-i} and suppose that, contrary to what we seek to show, there is ω_i, so that

$$\lim_{\epsilon \to 0^-} V(\varphi, \omega_i + \epsilon, \omega_{-i}) \quad \neq \quad \lim_{\epsilon \to 0^+} V(\varphi, \omega_i + \epsilon, \omega_{-i}) \ .$$

Specifically assume that

$$\lim_{\epsilon \to 0^-} V(\varphi, \omega_i + \epsilon, \omega_{-i}) \quad < \quad \lim_{\epsilon \to 0^+} V(\varphi, \omega_i + \epsilon, \omega_{-i}) \; .$$

Consider types $\omega_i' = \omega_i - \epsilon$ and $\omega_i'' = \omega_i + \epsilon$. If ω_i' makes a choices from the set of maximizers of type ω_i'' then this results in a payoff of

$$\hat{V} \quad := \quad V(\varphi, \omega_i'', \omega_{-i}) \; + \; k_i(y_i(\omega_i'', \cdot), \omega_i'') - k_i(y_i(\omega_i'', \cdot), \omega_i') \; ,$$

where, due to the continuity of the effort cost function,

$$\lim_{\omega_i' \to \omega_i''} \; k_i(y_i(\omega_i'', \cdot), \omega_i'') - k_i(y_i(\omega_i'', \cdot), \omega_i') \quad = \quad 0 \; .$$

Hence, for ϵ small enough, type ω_i' can make a choice that yields utility

$$\hat{V} \quad > \quad V(\varphi, \omega_i', \omega_{-i}) \; ,$$

which contradicts the assumption that the choice of type ω_i' is optimal. Thus, the assumption that the value function has a discontinuity has led to a contradiction and must be false.

A.1.3 Implementability with a continuous tax function

Consider the set of types with (ω, φ) with $\varphi \in \Phi^{12}(\omega)$ and let $I^{12}(\omega)$ be the indifference surface of type $\omega \in \mathbb{R}^2$,

$$I^{12}(\omega) = \Big\{ (c, y_1, y_2) \mid u(c) - k_1(y_1, \omega_1) - k_2(y_2, \omega_2) \; = \; v^{12}(\omega) \Big\} \; .$$

Note that any such surface can be described as a function $y \mapsto \mathcal{I}(y, \omega)$ with the interpretation, that given y_1 and y_2, c has to be equal to $\mathcal{I}(y, \omega)$ so as to keep type ω indifferent. Consider the lower envelope of the indifference surfaces for the different types ω, and note that this lower envelope is a continuous function $y \mapsto \underline{\mathcal{I}}(y)$. Now, construct a budget set $B^{12} \; = \; \{(c, y) \mid c = \underline{\mathcal{I}}(y)\}$. Choices from this budget set implement the allocation $y^*(\theta)$ conditional on $(\omega, \varphi) \in \Phi^{12}(\omega)$. We can now define

a corresponding tax function $\tilde{T}$ by letting

$$y_1 + y_2 - \tilde{T}(y) \quad = \quad \underline{\mathcal{I}}(y) \; .$$

or

$$\tilde{T}(y) \quad = \quad y_1 + y_2 - \underline{\mathcal{I}}(y) \; .$$

Upon letting $T(y) - b = \tilde{T}(y)$ and $T(0) = 0$ we obtain a continuous tax function T defined on the interior of $\mathcal{Y}$ so that $y^*(\theta)$ solves the problem

$$\max_{y \in \mathbb{R}^2} \quad u\Big(b + y_1 + y_2 - T(y)\Big) - k_1(y_1, \omega_1) - k_2(y_2, \omega_2) \; .$$

We proceed analogously for types with $(\omega, \varphi) \in \Phi^1(\omega)$ or $\Phi^2(\omega)$ and thereby obtain the values of T on the boundary of $\mathcal{Y}$.

A.2 Part (iii)

We show the following: If, for some l, k, ω_i' and ω_i'', $v_l(\omega_i'' \mid y_{-i}) > v_k(\omega_i'' \mid y_{-i})$ and $v_l(\omega_i' \mid y_{-i}) < v_k(\omega_i' \mid y_{-i})$, then there is $\hat{\omega}_i$ such that $v_k(\omega_i' \mid y_{-i}) > v_l(\omega_i' \mid y_{-i})$ for $\omega_i < \hat{\omega}_i$, and $v_k(\omega_i \mid y_{-i}) < v_l(\omega_i \mid y_{-i})$ for $\omega_i > \hat{\omega}_i$.

Note that, by the arguments in the proof of Part (i), both v_l and v_k are continuous functions. With

$$v_l(\omega_i'') - v_k(\omega_i'') \quad > \quad 0 \quad > \quad v_l(\omega_i') - v_k(\omega_i') \; ,$$

the intermediate value theorem implies that there exists a type $\hat{\omega}_i$ so that

$$v_l(\hat{\omega}_i) - v_k(\hat{\omega}_i) \quad = \quad 0 \; .$$

It remains to be shown that $\omega_i \mapsto v_l(\omega_i) - v_k(\omega_i)$ is a monotonic function. Suppose without loss of generality that, for given y_{-i}, $(y_1'', y_{-i}) \in P_l$ and $(y_1', y_{-i}) \in P_k$ imply that $y_1'' > y_1'$. Let $\underline{y}_{1l}^*$ be the infimum of $\operatorname{argmax}_{(y_i, y_{-i}) \in P_l} \; u(C(y_i, y_{-i})) - k_i(y_i, \omega_i)$. Let $\overline{y}_{1k}^*$ be the supremum of $\operatorname{argmax}_{(y_i, y_{-i}) \in P_k} \; u(C(y_i, y_{-i})) - k_i(y_i, \omega_i)$. Note that

$\underline{y}^*_{1l} \geq \overline{y}^*_{1k}$. By an application of Corollary 4 in Milgrom and Segal (2002), the derivative of $\omega_i \mapsto v_l(\omega_i) - v_k(\omega_i)$ with respect to ω_i is bounded from below by

$$-k_{\omega_i}(\underline{y}^*_{1l}, \omega_i) + k_{\omega_i}(\overline{y}^*_{1k}, \omega_i) \ .$$

This expression can be written as

$$-\int_{\underline{y}^*_{1l}}^{\overline{y}^*_{1k}} k_{y_i\omega_i}(s, \omega_i) ds$$

Since the cross derivative of the effort cost function is negative, it follows that $\omega_i \mapsto v_l(\omega_i) - v_k(\omega_i)$ is an increasing function.

A.3 Part (iv)

The extensive margin. By Part (ii), almost all types have a unique maximum over the set $\{U^{12}(\theta), U^1(\theta), U^2(\theta), U^0\}$. Put differently, with an atomless distribution, for any given ω, the types on the boundaries of the sets $\Phi^{12}(\omega)$, $\Phi^1(\omega)$ and $\Phi^2(\omega)$ have zero mass according to the probability measure $\tilde{F}^{\varphi}(\cdot \mid \omega)$.

The intensive margin. Consider the types for whom the optimal extensive margin choice is to pay both fixed costs and thus to have $y_1 > 0$ and $y_2 > 0$. Their intensive margin choice is to choose y_1 and y_2 to max

$$u(C(y_1, y_2)) - k_1(y_1, \omega_1) - k_2(y_2, \omega_2) \ . \tag{32}$$

The solution is a pair $(y_1^*(\omega), y_2^*(\omega))$ so that

$$y_1^*(\omega) \quad \in \quad \text{argmax}_{y_1} \quad u(C(y_1, y_2)) \quad - \quad k_1(y_1, \omega_1) \quad \text{s.t.} \quad y_2 = y_2^*(\omega) \ , \tag{33}$$

and

$$y_2^*(\omega) \quad \in \quad \text{argmax}_{y_2} \quad u(C(y_1, y_2)) \quad - \quad k_2(y_2, \omega_2) \quad \text{s.t.} \quad y_1 = y_1^*(\omega) \ . \tag{34}$$

By Part (iii) there are at most finitely many types in Ω_1, for which the subproblem in (33) has multiple solutions. Likewise, there are at most finitely many types in

Ω_2 for which the subproblem in (34) has multiple solutions. Thus, for almost all ω, the overall problem in (32) has a unique solution. The same argument can be applied to types for whom the optimal extensive margin choice is to pay only one of the fixed costs. Therefore, almost all types θ have a uniquely optimal intensive margin choice.

A.4 Part (v)

Optimal choices conditional on paying both fixed costs. Again, consider

$$\max_{y_1,y_2 \in Y^0} \; u(C(y_1,y_2)) - k_1(y_1,\omega_1) - k_2(y_2,\omega_2) \; .$$

We seek to show that, for almost all ω, the solution to this problem either satisfies first order conditions or strict bunching.

To do so, we consider the following class of problems: Treat y_2 as a fixed parameter. Choose y_1 to maximize $u(C(y_1,y_2)) - k_1(y_1,\omega_1)$. Again, we can decompose this problem in various subproblems, defined by the partition of $\mathcal{Y}_1$ that results from the restriction of $\mathcal{P}$ to income pairs with a fixed y_2, hence forth denoted by $\mathcal{P}_{y_2} = \{P_{1,y_2}, \ldots, P_{m,y_2}\}$.

Choices from some subset P_{j,y_2} of $\mathcal{Y}_1$ yield unique solutions due to the concavity of the budget frontier over the interior of the sets P_{j,y_2}. Multiple solutions can arise only when various such choices yield the same level of utility. In the following we show that, for any pair of subsets of P_{j,y_2}, P_{k,y_2}, (i) there are most two types in Ω_1 showing an indifference between their optimal choice from P_{j,y_2} and their optimal choice from P_{k,y_2}, (ii) all other types have a unique optimal choice, and this choice either satisfies a first order condition, or, is characterized by strict bunching.

Consider a level of income y_1' that lies at the boundary of some P_{j,y_2}. Assume that the marginal tax rate is discontinuous at (y_1', y_2), i.e.

$$\lim_{\epsilon \to 0^+} T_{y_1}(y_1' + \epsilon, y_2) \quad \neq \quad \lim_{\epsilon \to 0^-} T_{y_1}(y_1' + \epsilon, y_2).$$

Case 1: Suppose that the tax function is continuous at y_1', and that the marginal tax rate jumps upwards. Formally, $\lim_{\epsilon \to 0^+} T(y_1' + \epsilon, y_2) = \lim_{\epsilon \to 0^-} T(y_1' + \epsilon, y_2)$ and $\lim_{\epsilon \to 0^+} T_{y_1}(y_1' + \epsilon, y_2) > \lim_{\epsilon \to 0^-} T_{y_1}(y_1' + \epsilon, y_2)$.

Suppose there is a (possibly empty) subset of types $\Omega_1^B(y_1') \subset \Omega_1$ who choose y_1'. Given $c' = b + y_1' + y_2 - T(y_1', y_2)$, for all these types, the following inequalities both hold

$$u'(c')\Big(1 - \lim_{\epsilon \to 0^+} T_{y_1}(y_1' + \epsilon, y_2)\Big) \quad - \quad k_{y_1}(y_1', \omega_1) \quad \leq \quad 0 \tag{35}$$

and

$$u'(c')\Big(1 - \lim_{\epsilon \to 0^-} T_{y_1}(y_1' + \epsilon, y_2)\Big) \quad - \quad k_{y_1}(y_1', \omega_1) \quad \geq \quad 0 \tag{36}$$

so that there is neither an incentive to raise y_1 above y_1' nor an incentive to lower y_1 below y_1'. Let $\bar{\omega}_1$ be the level of ω_1 for which (35) holds as an equality. Let $\underline{\omega}_1$ be the level of ω_1 for which (36) holds as an equality. The assumptions that define Case 1 imply that $\underline{\omega}_1 < \bar{\omega}_1$, see Figure 2.

Now suppose that $\Omega_1^B(y_1')$ contains types (ω_1, ω_2) so that $\omega_1 \in]\underline{\omega}_1, \bar{\omega}_1[$. For such a type, conditions (35) and (36) hold as strict inequalities. Thus, there are at most two types in $\Omega_1^B(y_2)$ that do not satisfy the condition of strict bunching. As the figure shows, for all types in $\Omega_1^B(y_1')$, y_1' is the unique solution to the problem of maximizing utility given y_2.

Case 2: Suppose that the tax function is continuous at y_1', and that the marginal tax rate jumps downwards. Formally, $\lim_{h \to 0^+} T(y_1' + h, y_2) = \lim_{h \to 0^-} T(y_1' + h, y_2)$ and $\lim_{h \to 0^+} T_{y_1}(y_1' + h, y_2) < \lim_{h \to 0^-} T_{y_1}(y_1' + h, y_2)$. This creates a local convexity in the budget set, with the implication that choosing y_1' is dominated. Thus, $\Omega_1^B(y_1') = \emptyset$, see Figure 2.

Completing the proof requires to deal with further case distinctions that involve notches rather than kinks. This is somewhat repetitive. The arguments are the same as the ones in Cases 1 and 2. We relegate these steps to the Supplementary material for this paper.

B Proof of Proposition 2

B.1 Differentiability of tax revenue

Lemma 1 *The function Δ^R is differentiable.*

Proof. By Proposition 1, almost all types have unique optimal earnings choices. We now argue that those choices depend in a differentiable way on the parameters that define the reform (s, h).

If the optimal choice of y_i is characterized by strict bunching, then a marginal change of any one parameter will not affect this optimal choice, implying a derivative of zero. If the optimal choices of y_i and y_{-i} belong to a box B^{jk} and satisfy first order conditions, then they solve a system of two equations and, given our assumptions on preferences and the tax function, the solution depends, for almost all types and almost all levels income, in a differentiable way, on the parameters that define (s, h). Specifically, the derivatives with respect the parameters that matter locally – δ and $s\, b^{jk}$, $s\, \tau_1^{jk}$, $s\, \tau_2^{jk}$ – will generally be different from zero, while the derivatives with respect to parameters that change the tax function for y outside of B^{jk} are equal to zero. Likewise, when y_i satisfies a first order condition and y_{-i} is characterized by strict bunching, then, y_i solves one first order condition, for fixed y_{-i}. Again, the solution depends in a differentiable way on the parameters that define (s, h). Thus, in particular, for almost all ω, and $x \in \{\delta, (s\, b^{jk}, s\, \tau_1^{jk}, s\, \tau_2^{jk})\}$ the derivatives

$$\frac{d}{dx}\ T^n(y_{1d}^*(\cdot), y_{2d}^*(\cdot)) \quad = \quad T_1^n(\cdot)\ \frac{dy_{1d}^*(\cdot)}{dx}\ +\ T_2^n(\cdot)\ \frac{dy_{2d}^*(\cdot)}{dx}$$

$$\frac{d}{dx}\ T^n(y_{1s}^*(\cdot), 0) \quad = \quad T_1^n(\cdot)\ \frac{dy_{1s}^*(\cdot)}{dx}$$

and

$$\frac{d}{dx}\ T^n(0, y_{2s}^*(\cdot)) \quad = \quad T_2^n(\cdot)\ \frac{dy_{2s}^*(\cdot)}{dx}$$

are well-defined. These expressions capture how the tax payments of individuals with type ω are affected by behavioral responses at the intensive margin. Extensive margin responses are captured by the derivatives of $\sigma^{12}(s, h, \omega)$, $\sigma^1(s, h, \omega)$ and $\sigma^2(s, h, \omega)$. Again, it follows from Proposition 1, that these derivatives are well

defined for almost all ω. Consider e.g.

$$\sigma^{12}(s,h,\omega) \quad = \quad \mathbf{E}_{\varphi}[\mathbf{1}(\varphi \in \Phi^{12}(s,h,\omega)) \mid \omega]$$

and note that

$$\sigma^{12}(s,h,\omega) = \int_0^{v^{12}(\cdot)-v^2(\cdot)} \int_0^{\bar{\varphi}_2(s,h,\omega)} \tilde{f}(\varphi \mid \omega)\, d\varphi_2\, d\varphi_1 \ , \tag{37}$$

where $\bar{\varphi}_2(s,h,\omega) = \min\{v^{12}(\cdot) - v^0(\cdot) - \varphi_1, v^{12}(\cdot) - v^1(\cdot)\}$. By the arguments in the proof of Proposition 1, for almost all types, the subproblems that yield the value functions $v^{12}(\cdot)$, $v^1(\cdot)$ and $v^2(\cdot)$ have unique solutions. By Corollary 4 in Milgrom and Siegel (2002) they are therefore differentiable functions of ω, and of all x in $\{\delta, (s\ b^{jk}, s\ \tau_1^{jk}, s\ \tau_2^{jk})\}$. Consequently, the expression on the right hand side of (37) is a differentiable function of any such x, for almost all ω. The partial derivative of $\sigma^{12}(\cdot)$ with respect to any such x can therefore be computed using a repeated application of Leibnitz' rule. Likewise, for $\sigma^1(\cdot)$ and $\sigma^2(\cdot)$. These derivatives are measures of behavioral responses at the extensive margin.Thus, for almost all ω, all the expressions that appear on the right hand side of (9) are differentiable functions of any x in $\{\delta, (s\ b^{jk}, s\ \tau_1^{jk}, s\ \tau_2^{jk})\}$. Therefore, $\mathbf{E}_{\theta}[T^n(y^*(s,h,\theta))]$ is differentiable with respect to any such x.

B.2 Characterization of $R_s(0,h)$.

Budget balance requires the existence of δ solving the fixed point equation,

$$\delta \quad = \quad \Delta(\delta,\cdot) - R^0 \ . \tag{38}$$

With δ bounded from above, the existence of a fixed point then follows from Brouwer's fixed point theorem upon noting that the function $\delta \mapsto \Delta^R(\delta,\cdot)$ is continuous.[56] When T is a non-decreasing function, uniqueness is established as follows: With concave consumption utility, indifference curves in both the c-y_1 and in the c-y_2 space become steeper when δ increases. Thus, individuals choose (weakly) lower incomes when δ increases. With T non-decreasing, it follows that $\delta \mapsto \Delta^R(\delta,\cdot)$ is

56. Continuity is implied by differentiability which has been established in Lemma 1.

a decreasing function. Thus, the left hand side of Equation (38) is increasing in δ and the right-hand side is decreasing, with the implication that they intersect at most once.

We denote the fixed point by $R(s,h)$ and note that it is implicitly defined by

$$R(s,h) \;=\; \Delta(R(s,h),\cdot) \;-\; R^0 \,.$$

Lemma 1 now enables to apply the implicit function theorem and to compute the partial derivative of the function R with respect to any $x \in \{s\, b^{jk}, s\, \tau_1^{jk}, s\, \tau_2^{jk}\}$ as

$$R_x(s,h) \;=\; \frac{\Delta_x\Big(\delta, \{s\, b^{jk}, s\, \tau_1^{jk}, s\, \tau_2^{jk}\}\Big)}{1-\Delta_\delta\Big(\delta, \{s\, b^{jk}, s\, \tau_1^{jk}, s\, \tau_2^{jk}\}\Big)} \,.$$

At the status quo, i.e. for $s=0$, we have

$$R_x(0,h) \;=\; \frac{\Delta_x^0}{1-\Delta_\delta^0} \,. \tag{39}$$

Changing the scale s. The marginal effect on tax revenue, starting from the status quo is given by

$$\frac{d}{ds}\, R(s,h)_{|s=0} \;=:\; R_s(0,h)$$

For arbitrary s, an application of the chain rule yields

$$\frac{d}{ds}\, R(s,h) \;=\; \frac{1}{1-\Delta_\delta(\cdot)} \sum_{(j,k)} \Big(b^{jk}\Delta_{sb^{jk}}(\cdot) + \tau_1^{jk}\Delta_{s\tau_1^{jk}}(\cdot) + \tau_2^{jk}\Delta_{s\tau_2^{jk}}(\cdot)\Big)$$

With the more concise notation introduced in the main text, this can be written as

$$\frac{d}{ds}\, R(s,h) \;=\; \frac{1}{1-\Delta_\delta(\cdot)} \sum_{(j,k)} \Big(b^{jk}\Delta_b(\cdot \mid y^{jk},\ell) + \tau_1^{jk}\Delta_1(\cdot \mid y^{jk},\ell) + \tau_2^{jk}\Delta_2(\cdot \mid y^{jk},\ell)\Big)$$

At the status quo this becomes

$$R_s(0,h) \;=\; \frac{1}{1-\Delta_\delta^0} \sum_{(j,k)} \Big(b^{jk}\Delta_b^0(y^{jk},\ell) + \tau_1^{jk}\Delta_1^0(y^{jk},\ell) + \tau_2^{jk}\Delta_2^0(y^{jk},\ell)\Big) \,.$$

B.3 Characterization of $\Delta_x^0(y^{jk}, \ell)$

B.3.1 Derivation of Equation (10)

Suppose, without loss of generality, that there is only one box B^{jk} with $b^{jk} \neq 0$, $\tau_1^{jk} \neq 0$ or $\tau_2^{jk} \neq 0$. Let

$$\Omega^{jk}(s,h) \quad = \quad \{\omega \mid y^{d*}(\omega, s, h) \in B^{jk}\} \ .$$

Let

$$\underline{\omega}_1(s,h) \quad = \quad \min\{\omega_1 \mid \exists \omega_2 \text{ s.t } (\omega_1, \omega_2) \in \Omega^{jk}(s,h)\}$$

and

$$\overline{\omega}_1(s,h) \quad = \quad \max\{\omega_1 \mid \exists \omega_2 \text{ s.t } (\omega_1, \omega_2) \in \Omega^{jk}(s,h)\} \ .$$

For given $\omega_1 \in \Omega_1^{jk}(\tau, h) := [\underline{\omega}_1(s,h), \overline{\omega}_1(s,h)]$, let

$$\underline{\omega}_2(\omega_1, s, h) \quad = \quad \min\{\omega_2 \mid (\omega_1, \omega_2) \in \Omega^{jk}(s,h)\}$$

and

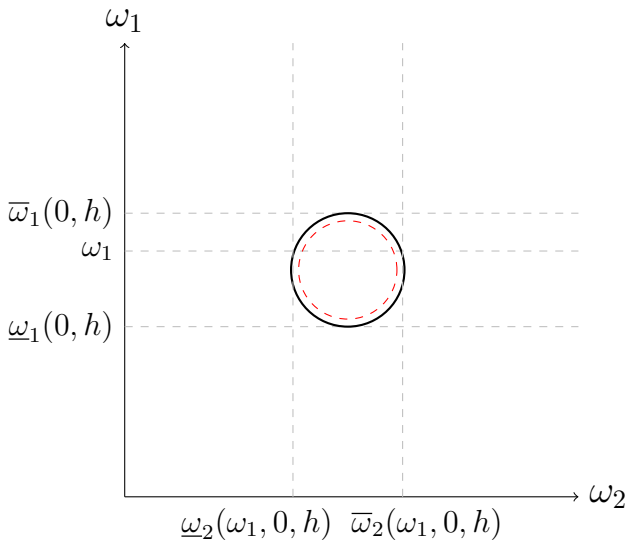
$$\overline{\omega}_2(\omega_1, s, h) \quad = \quad \max\{\omega_2 \mid (\omega_1, \omega_2) \in \Omega^{jk}(s,h)\} \ .$$

Armed with this notation we can write

$$\Delta(\cdot) \ = \ \mathbf{E}_\omega[\big(r(\omega, s, h) - r^0(\omega)\big)\mathbf{1}(\omega \in \Omega^{jk}(s,h))]$$

$$\int_{\underline{\omega}_1(s,h)}^{\overline{\omega}_1(s,h)} \int_{\underline{\omega}_2(\omega_1,\tau,h)}^{\overline{\omega}_2(\omega_1,s,h)} \big(r(\omega, s, h) - r^0(\omega)\big) \tilde{f}_{\omega_2}(\omega_2 \mid \omega_1) d\omega_2 \ \tilde{f}_{\omega_1}(\omega_1) d\omega_1 \ ,$$

where $\tilde{f}_{\omega_1}$ is the marginal distribution of ω_1 associated with $\tilde{F}$ and $\tilde{f}_{\omega_2}(\cdot \mid \omega_1)$ is the distribution of ω_2 conditional on ω_1. Differentiating with respect to some



Note: The area in the black circle is the set $\Omega^{jk}(0,h)$. (The circle is a generic shape. There is no presumption that the set takes this specific geometric form.) Suppose a reform implies that box B^{jk} becomes less attractive and that the set $\Omega^{jk}(s,h)$ is therefore decreasing in s, as indicated by the dashed red circle, which represents $\Omega^{jk}(s,h)$ for some positive level of s. At $s=0$, the revenue loss from the fact that fewer agents make choices in the box B^{jk} is equal to zero. The reason is that every loss of an agent with a type ω on the black circle is multiplied with the change in their tax payment. At $s=0$, this change is zero, $r(\omega,0,h)=0$.

$x \in \{sb^{jk}, s\tau_1^{jk}, s\tau_2 jk\}$ and repeatedly applying of Leibnitz' rule yields

$$\Delta_x(\cdot) \;=\; \int_{\underline{\omega}_1(s,h)}^{\overline{\omega}_1(s,h)} \int_{\underline{\omega}_2(\omega_1,s,h)}^{\overline{\omega}_2(\omega_1,s,h)} r_x(\omega,s,h)\; \tilde{f}_{\omega_2}(\omega_2 \mid \omega_1) d\omega_2 \; \tilde{f}_{\omega_1}(\omega_1) d\omega_1 \;,$$

$$+\; A(s,h) \;+\; B(s,h) \;+\; C(s,h) \;+\; D(s,h) \;,$$

where

$$A(s,h) \;=\; \left(\int_{\underline{\omega}_2(\overline{\omega}_1(\cdot),s,h)}^{\overline{\omega}_2(\overline{\omega}_1(\cdot),s,h)} \left(r(\overline{\omega}_1(\cdot),\omega_2,s,h) - r^0(\overline{\omega}_1(\cdot),\omega_2) \right) \tilde{f}_{\omega_2}(\omega_2 \mid \omega_1) d\omega_2 \right) \; \frac{\partial}{\partial x} \overline{\omega}_1(s,h) \;,$$

$$B(s,h) \;=\; -\left(\int_{\underline{\omega}_2(\underline{\omega}_1(\cdot),s,h)}^{\overline{\omega}_2(\underline{\omega}_1(\cdot),s,h)} \left(r(\underline{\omega}_1(\cdot),\omega_2,s,h) - r^0(\underline{\omega}_1(\cdot),\omega_2) \right) \tilde{f}_{\omega_2}(\omega_2 \mid \omega_1) d\omega_2 \right) \; \frac{\partial}{\partial x} \underline{\omega}_1(s,h) \;,$$

$$C(s,h) \;=\; \int_{\underline{\omega}_1(s,h)}^{\overline{\omega}_1(s,h)} \left(r(\omega_1,\overline{\omega}_2(\cdot),s,h) - r^0(\omega_1,\overline{\omega}_2(\cdot)) \right) \; \frac{\partial}{\partial x} \underline{\omega}_2(\omega_1,s,h) \; \tilde{f}_{\omega_1}(\omega_1) d\omega_1 \;,$$

and

$$D(s,h) \;=\; -\int\limits_{\underline{\omega}_1(s,h)}^{\overline{\omega}_1(s,h)} \left(r(\omega_1,\underline{\omega}_2(\cdot),s,h) - r^0(\omega_1,\underline{\omega}_2(\cdot))\right) \tfrac{\partial}{\partial x}\underline{\omega}_2(\omega_1,s,h)\; \tilde{f}_{\omega_1}(\omega_1)d\omega_1 \;.$$

Note that $s=0$ implies that $T^n(y)=T(y)$, for all y. Hence $r(\omega,0,h)=r^0(\omega)$, for all ω and all h. This implies, moreover, that

$$A(0,h) \;=\; B(0,h) \;=\; C(0,h) \;=\; D(0,h) \;=\; 0\;.$$

Therefore,

$$\Delta_x^0 \;=\; \int\limits_{\underline{\omega}_1(0,h)}^{\overline{\omega}_1(0,h)} \int\limits_{\underline{\omega}_2(\omega_1,0,h)}^{\overline{\omega}_2(\omega_1,0,h)} r_x(\omega,0,h)\; \tilde{f}_{\omega_2}(\omega_2 \mid \omega_1)d\omega_2\; \tilde{f}_{\omega_1}(\omega_1)d\omega_1 \;,$$

$$\;=\; \mathbb{E}_\omega[r_x(\omega,0,h)\; \mathbf{1}(\omega \in \Omega^{jk}(0,h))]\;.$$

To complete the derivation of (10) we need to show that, for all $x \in \{sb^{jk}, s\tau_1^{jk}, s\tau_2^{jk}\}$,

$$r_x(\omega,0,h) \;=\; \beta_{I,x}^0(\omega) + \beta_{E,x}^0(\omega) + \mu_x^0(\omega)\;,$$

showing that every change of a local tax parameter x has revenue implications that can be written as a sum of a mechanical effect and of revenue effects from behavioral responses at the intensive and the extensive margin. This part of the proof involves only routine steps: Computing derivatives and collecting terms. It is relegated to part G of the Supplementary Material.

C Proof of Proposition 3

We detail the main steps in the derivation of Equation (13), with some more technical steps in this derivation relegated to Section H of the Supplementary material. There, we also analyze the special case with $u(c)=c$ and show that, under this additional assumption, $\Delta_\delta^0=0$ and, for all y, $\bar{\beta}_{I,b}^0(y)=0$.

Let h be a continuous function. Consider a grid of $\mathcal{Y}$ and the boxes B^{jk} with lower left corners y^{jk}. Now define an approximation h^n of h in the following way: For each B^{jk}, let $b^{jk}=h(y^{jk})=h^n(y^{jk})$, which implies, in particular, that the

functions h and h^n coincide at the corner of the boxes. Moreover, for $y \in B^{jk}$, let $h^n(y) = h^n(y^{jk}) + \tau_1^{jk}(y_1 - y_1^{jk}) + \tau_2^{jk}(y_2 - y_2^{jk})$, where $\tau_1^{jk} = \frac{h(y^{(j+1)k}) - h(y^{jk})}{\ell_1}$ and $\tau_2^{jk} = \frac{h(y^{(j+1)(k+1)}) - h(y^{(j+1)k})}{\ell_2}$. Thus, h^n is a piecewise linear approximation of h. For later reference, note that $\ell_1 \to 0$ implies that $\tau_1^{jk} \to h_1(y^{jk})$, where h_1 is the derivative of the function h with respect to its first argument.[57] Likewise, $\ell_2 \to 0$ implies $\tau_2^{jk} \to h_2(y^{jk})$. By Proposition 2,

$$(1 - \Delta_\delta^0) R_s(0, h) \;\; := \;\; \sum_{(j,k)} \Big(b^{jk} \Delta_b^0(y^{jk}, \ell) + \tau_1^{jk} \Delta_1^0(y^{jk}, \ell) + \tau_2^{jk} \Delta_2^0(y^{jk}, \ell) \Big) \;, \quad (40)$$

where, for any $x \in \{s b^{jk}, s\tau_1^{jk}, s\tau_2^{jk}\}$,

$$\Delta_x^0(y^{jk}, \ell_1, \ell_2) \;\; := \;\; \mathbf{E}_\omega \left[\mathbf{1}\left(y_d^0(\omega) \in B^{jk} \right) r_x^0(\omega) \right] \;.$$

and

$$r_x(\omega, 0, h) \quad = \quad \beta_{I,x}^0(\omega) + \beta_{E,x}^0(\omega) + \mu_x^0(\omega) \;,$$

The revenue effects collected in $\Delta_x^0(y^{jk}, \ell)$ are a convex combination of revenue effects from individuals who bunch at a mass point in B^{jk} and due to individuals who also have earnings in B^{jk}, but are not bunching. To make this explicit, we use the following notation: Let $Y_M^{jk} = \{y_m^{jk}\}$, for $m \in \{1, \ldots, M^{jk}\}$ be the collection of mass points in B^{jk}. Let Y_M be the set of all mass points and note that the total number of mass points is finite. Denote by $f^M(y_m^{jk})$ the mass associated with $y_m^{jk} \in Y_M^{jk}$. Denote by

$$\nu(y^{jk}, \ell) := \mathbf{E}_\omega \left[\mathbf{1}\left(y_d^0(\omega) \in B^{jk} \right) \right] - \bar{f}_{jk}^M$$

the complementary atomless mass in box B^{jk}, where $\bar{f}_{jk}^M := \sum_{m=1}^{M^{jk}} f^M(y_m^{jk})$. Equation (10) can now be written as

$$\Delta_x^0(y^{jk}, \ell) \quad = \quad \sum_{m=1}^{M^{jk}} f^M(y_m^{jk}) \; \bar{r}_x^0(y_m^{jk}) \quad + \quad \nu(y^{jk}, \ell) \; \bar{r}_x^0(y^{jk}, \ell) \;, \quad (41)$$

where $\bar{r}_x^0(y_m^{jk}) = \mathbf{E}_\Omega \left[r_x^0(\omega) \mid y_d^0(\omega) = y_m^{jk} \right]$ is the average effect from individuals at

57. h is a continuous function on a bounded domain and hence differentiable almost everywhere.

mass point y_m^{jk} and $\bar{r}_x^0(y^{jk},\ell) := \mathbf{E}_\Omega\left[r_x^0(\omega) \mid y_d^0(\omega) \in B^{jk} \setminus Y_M^{jk}\right]$ is the average effect from individuals who are not at a mass point. We can now write

$$\begin{aligned}(1-\Delta_\delta^0)R_s(0,h^n) &:= \sum_{(j,k)}\Big(h^n(y^{jk})\Delta_b^{0\nu}(y^{jk},\ell)+\tau_1^{jk}\Delta_1^{0\nu}(y^{jk},\ell)+\tau_2^{jk}\Delta_2^{0\nu}(y^{jk},\ell)\Big)\\ &\quad+\sum_{(j,k)}\sum_{y_m^{jk}\in Y_M} h^n(y_m^{jk})\;\mathbf{1}(y_m^{jk}\in B^{jk})\;f^M(y_m^{jk})\;\bar{r}_b^0(y_m^{jk})\end{aligned} \tag{42}$$

where, for any $x \in \{sb^{jk}, s\tau_1^{jk}, s\tau_2^{jk}\}$,

$$\Delta_x^{0\nu}(y^{jk},\ell_1,\ell_2) := \nu(y^{jk},\ell)\;\bar{r}_x^0(y^{jk},\ell)\;,$$

where we have been using the insight from part H.1 of the Supplement that, for any mass point y_m,

$$\tau_1^{jk}\;\mathbf{1}(y_m\in B^{jk})\;\bar{r}_1^0(y_m)+\tau_1^{jk}\;\mathbf{1}(y_m\in B^{jk})\;\bar{r}_2^0(y_m)=0\;,$$

since bunching taxpayers show no intensive margin response.

Making the grid finer. Consider a sequence of grids, indexed by n. Let $\ell_1^n = \frac{1}{2}\ell_1^{n-1}$ and $\ell_2^n = \frac{1}{2}\ell_2^{n-1}$. Hence, the boxes in partition n have half the length of the boxes in partition $n-1$. By construction, if y^{jk} is the lower left corner of some box in partition n, then it is also the lower left corner of a box in any partition $n' > n$.

Step 1. By construction of the sequence of grids, as $n\to\infty$, $\mid y_m - y_n^{jk} \mid \to 0$ where y_n^{jk} is the lower left corner of the box that contains y_m. Since h is continuous this also implies that $\mid h(y_m) - h^n(y_n^{jk}) \mid = \mid h(y_m) - h(y_n^{jk}) \mid \to 0$. Therefore,

$$\sum_{(j,k)}\sum_{y_m^{jk}\in Y_M} h^n(y_m^{jk})\;\mathbf{1}(y_m^{jk}\in B^{jk})\;f^M(y_m^{jk})\;\bar{r}_b^0(y^j k_m)$$

converges to

$$\sum_{y_m\in Y_M} h(y_m)\bar{\mathcal{R}}_b^M(y)\;,$$

as $n\to\infty$, where $\bar{R}_b^M(y) = f^M(y)\;\bar{r}_b^0(y)$ for $y\in Y^M$.

Step 2. Show that $\sum_{(j,k)} \Big(h^n(y^{jk}) \Delta_b^{0\nu}(y^{jk}, \ell^n) + \tau_1^{jk} \Delta_1^{0\nu}(y^{jk}, \ell^n) + \tau_2^{jk} \Delta_2^{0\nu}(y^{jk}, \ell^n) \Big)$ converges to $\int_{\mathcal{Y}} \Big(h(y)\ \bar{\mathcal{R}}_b^A(y) + h_1(y)\ \bar{\mathcal{R}}_1(y) + h_2(y)\ \bar{\mathcal{R}}_2(y) \Big) dy$ as $n \to \infty$.

Assume in the following that expressions such as $\Delta_x^{0\nu}(y^{jk}, \ell_1^n, \ell_2^n)$ are twice differentiable functions of ℓ_1 and ℓ_2. This assumption is verified in Section H of the Supplementary material . There we also show that, with ν an atomless distribution, for ℓ_1 and ℓ_2 close to zero, first order Taylor approximations imply

$$\Delta_x^{0\nu}(y^{jk}, \ell_1^n, \ell_2^n) \quad \simeq \quad \ell_1\ \ell_2\ \Delta_{x,\ell_2\ell_1}^{0\nu}(y^{jk}, 0, 0)\ , \tag{43}$$

where the subscript $\ell_2\ell_1$ indicates a cross-derivative. In Section H.2 of the Supplementary material, we also verify that

$$\bar{\mathcal{R}}_x(y^{jk}) = \Delta_{x,\ell_2\ell_1}^{0\nu}(y^{jk}, 0, 0) = \bar{r}_x(y^{jk}) f^A(y^{jk})\ ,$$

where f^A is a density over the atomless part of the income space. Thus, for ℓ_1^n and ℓ_2^n close to zero,

$$\begin{array}{l} \sum_{(j,k)} \Big(h^n(y^{jk}) \Delta_b^{0\nu}(y^{jk}, \ell^n) + \tau_1^{jk} \Delta_1^{0\nu}(y^{jk}, \ell^n) + \tau_2^{jk} \Delta_2^{0\nu}(y^{jk}, \ell^n) \Big) \\ \simeq \quad \sum_{(j,k)} \ell_1^n \ell_2^n \Big(h^n(y^{jk})\ \bar{\mathcal{R}}_b(y^{jk}) + h_1^n(y^{jk})\ \bar{\mathcal{R}}_1(y^{jk}) + h_2^n(y^{jk})\ \bar{\mathcal{R}}_2(y^{jk}) \Big)\ . \end{array} \tag{44}$$

where $h_1^n(y^{jk}) = \tau_1^{jk}$ is the derivative of the function h^n with respect to its first argument for $y \in B^{jk}$; analogously $h_2^n(y^{jk}) = \tau_2^{jk}$. As $n \to \infty$ the difference between the left hand side and the right hand side of (44) vanishes. Note that $\mid h(y^{jk}) - h^n(y^{jk}) \mid$, $\mid h_1(y^{jk}) - h_1^n(y^{jk}) \mid$ and $\mid h_2(y^{jk}) - h_2^n(y^{jk}) \mid$ all go to zero as $\ell_1^n \to 0$ and $\ell_2^n \to 0$. Therefore, also

$$\begin{array}{l} \sum_{(j,k)} \Big(h^n(y^{jk}) \Delta_b^{0\nu}(y^{jk}, \ell^n) + \tau_1^{jk} \Delta_1^{0\nu}(y^{jk}, \ell^n) + \tau_2^{jk} \Delta_2^{0\nu}(y^{jk}, \ell^n) \Big) \\ \simeq \quad \sum_{(j,k)} \ell_1^n \ell_2^n \Big(h(y^{jk})\ \bar{\mathcal{R}}_b(y^{jk}) + h_1(y^{jk})\ \bar{\mathcal{R}}_1(y^{jk}) + h_2(y^{jk})\ \bar{\mathcal{R}}_2(y^{jk}) \Big)\ . \end{array} \tag{45}$$

Again, as $n \to \infty$ the difference between the left hand side and the right hand side of (45) vanishes. The expression on the right hand-side is a Riemann sum that converges to the corresponding Riemann integrals as $n \to \infty$ and hence $\ell_1^n \to 0$ and

$\ell_2^n \to 0$. Thus, as $n \to \infty$,

$$\sum_{(j,k)} \Big(h^n(y^{jk}) \Delta_b^{0\nu}(y^{jk}, \ell^n) + \tau_1^{jk} \Delta_1^{0\nu}(y^{jk}, \ell^n) + \tau_2^{jk} \Delta_2^{0\nu}(y^{jk}, \ell^n) \Big)$$

converges to

$$\int_{\mathcal{Y}} \Big(h(y)\ \bar{\mathcal{R}}_b^A(y) + h_1(y)\ \bar{\mathcal{R}}_1(y) + h_2(y)\ \bar{\mathcal{R}}_2(y) \Big) dy\ . \tag{46}$$

with $\bar{\mathcal{R}}_b^A(y) := \bar{r}_b(y) f^A(y)$, $\bar{\mathcal{R}}_1(y) := \bar{r}_b(y) f^A(y)$ and $\bar{\mathcal{R}}_2(y) := \bar{r}_b(y) f^A(y)$.

Step 3. Steps 1 and 2 imply that the right-hand side of (42) converges to

$$\begin{array}{l} \int_{\mathcal{Y}} \Big(h(y)\ \bar{\mathcal{R}}_b^A(y) + h_1(y)\ \bar{\mathcal{R}}_1(y) + h_2(y)\ \bar{\mathcal{R}}_2(y) \Big) dy\ . \\ \quad + \sum_{y_m \in Y_M} h(y_m) \bar{\mathcal{R}}_b^M(y_m). \end{array} \tag{47}$$

Define $\bar{\mathcal{R}}_b(y) := \bar{\mathcal{R}}_b^A(y) + \mathbf{1}(y \in Y_M) \bar{\mathcal{R}}_b^M(y)$, and analogously for $\bar{\mathcal{R}}_1(y)$ and $\bar{\mathcal{R}}_2(y)$. This allows us to write

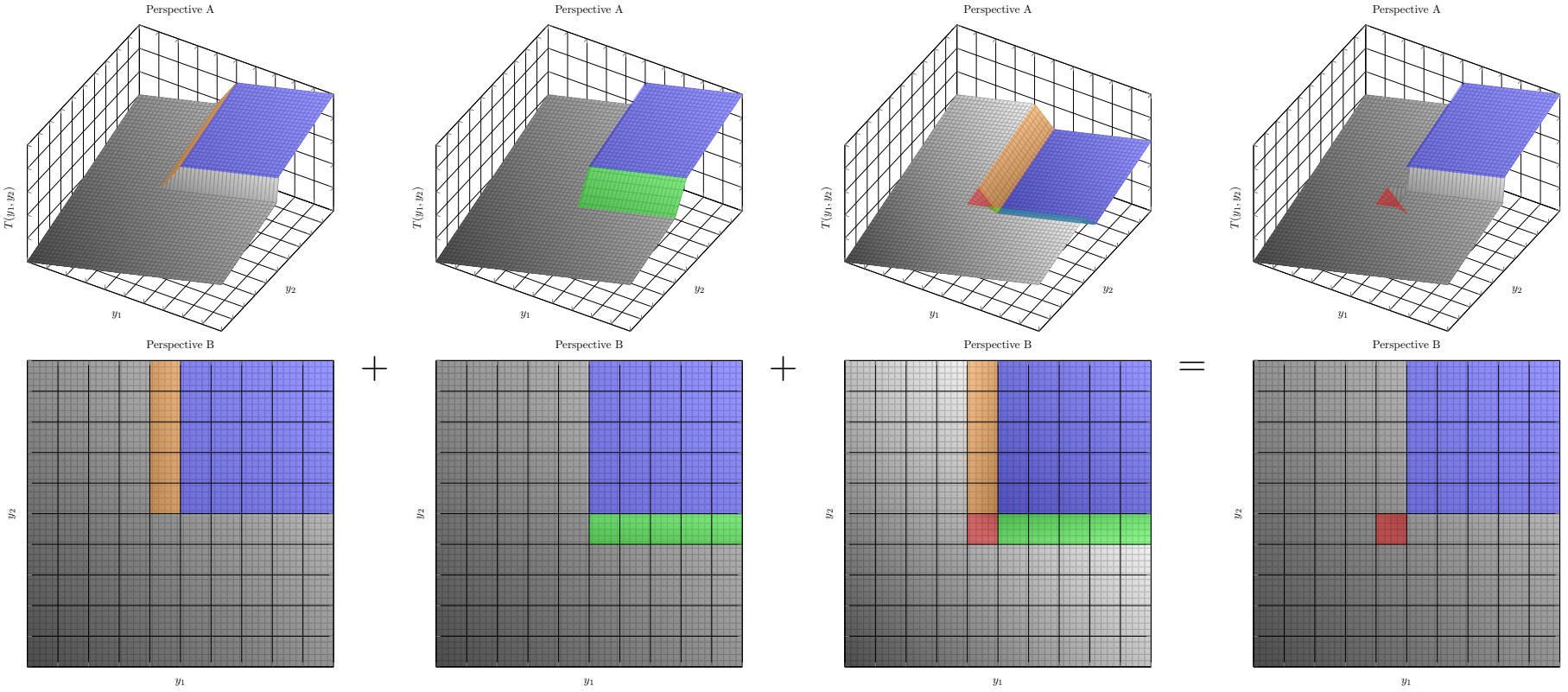
$$\int_{\mathcal{Y}} h(y)\ \bar{\mathcal{R}}_b(y) dy \ \ := \ \ \int_{\mathcal{Y}} h(y)\ \bar{\mathcal{R}}_b^A(y) dy + \int_{\mathcal{Y}} \mathbf{1}(y \in Y^M)\ h(y)\ \bar{\mathcal{R}}_b^M(y) dy$$

rather than $\int_{\mathcal{Y}} h(y)\ \bar{\mathcal{R}}_b^A(y) dy + \sum_{y_m \in Y_M} h(y_m) \bar{\mathcal{R}}_b^M(y_m)$. Therefore, we can write the expression in (47) also as

$$\int_{\mathcal{Y}} \Big(h(y)\ \bar{\mathcal{R}}_b(y) + h_1(y)\ \bar{\mathcal{R}}_1(y) + h_2(y)\ \bar{\mathcal{R}}_2(y) \Big) dy\ .$$

This completes the derivation of Equation (13). The claims on the implications of $u(c) = c$ are proven in Part H of the Supplement.

Figure C1: Illustration of the Impossibility Theorem



D Proof of Proposition 4

D.1 Preliminaries

In Part G.2 of the Supplement, we formally derive all the partial derivatives that are needed to characterize the revenue effects of a tax reform. This decomposition into behavioral responses and mechanical effects implies that, in the absence of extensive margin responses,

$$\bar{\beta}^0_{I,1}(y) \quad = \quad T_1(y^0_d(\bar{\omega}(y)))\; y^0_{d1,1}(\bar{\omega}(y)) + T_2(y^0_d(\bar{\omega}(y)))y^0_{d2,1}(\bar{\omega}(y))$$

and

$$\bar{\beta}^0_{I,2}(y) \quad = \quad T_1(y^0_d(\bar{\omega}(y)))\; y^0_{d1,2}(\bar{\omega}(y)) + T_2(y^0_d(\bar{\omega}(y)))y^0_{d2,2}(\bar{\omega}(y))$$

Moreover,

$$-y^*_{d1,x}(\cdot)\,[T_{11}(\cdot) + k_{1,yy}(\cdot)] - \; y^*_{d2,x}(\cdot)\;T_{12}(\cdot) - \; \mathbf{1}(x = \bar{\tau}^{jk}_1) \;=\; 0\;, \tag{48}$$

and

$$-y^*_{d2,x}(\cdot)\,[\;T_{22}(\cdot) + k_{2,yy}(\cdot)] - \; y^*_{d1,x}(\cdot)\;T_{21}(\cdot) - \; \mathbf{1}(x = \bar{\tau}^{jk}_2) \;=\; 0\;. \tag{49}$$

Imposing that $\mathbf{1}(x = \bar{\tau}^{jk}_1) = 1$ and $\mathbf{1}(x = \bar{\tau}^{jk}_2) = 0$ yields

$$y^0_{d1,1} \quad = \quad -\frac{T_{22} + k_{2,yy}}{det(H)}\;, \tag{50}$$

and

$$y^0_{d2,1} \quad = \quad \frac{T_{12}}{det(H)}\;, \tag{51}$$

where

$$det(H) \quad = \quad [T_{11}(\cdot) + k_{1,yy}(\cdot)]\,[\;T_{22}(\cdot) + k_{2,yy}(\cdot)] - [T_{12}]^2$$

is the determinant of the Hessian matrix that follows from the objective of the utility-maximization problem. For an interior solution that is characterized by first order conditions we need to have $det(H) > 0$. Analogously, we obtain

$$y^0_{d1,2} \quad = \quad \frac{T_{12}}{det(H)} \, , \tag{52}$$

and

$$y^0_{d2,2} \quad = \quad -\frac{T_{11} + k_{1,yy}}{det(H)} \, , \tag{53}$$

Plugging these expressions into the formula for $\bar{\beta}^0_{I,1}(y)$ yields

$$\bar{\beta}^0_{I,1}(y) = -\frac{1}{det(H)} \left[T_1(\cdot)(T_{22}(\cdot) + k_{2,yy}(\cdot)) - T_2(\cdot) \, T_{12}(\cdot) \right] \tag{54}$$

and analogously,

$$\bar{\beta}^0_{I,2}(y) = -\frac{1}{det(H)} \left[T_2(\cdot)(T_{11}(\cdot) + k_{1,yy}(\cdot)) - T_1(\cdot) \, T_{12}(\cdot) \right] . \tag{55}$$

Iso-elastic effort cost functions. Suppose that the variable effort cost functions k_1 and k_2 are iso-elastic; i.e.

$$k_1(y_1, \omega_1) \quad = \quad \frac{1}{1 + \frac{1}{\varepsilon_1}} \left(\frac{y_1}{\omega_1} \right)^{1 + \frac{1}{\varepsilon_1}} \quad \text{and} \quad k_2(y_2, \omega_2) \quad = \quad \frac{1}{1 + \frac{1}{\varepsilon_2}} \left(\frac{y_2}{\omega_2} \right)^{1 + \frac{1}{\varepsilon_2}} \, .$$

Then, we have that, for any $i = 1, 2$,

$$k_{i,yy}(\cdot) \quad = \quad \frac{1}{\varepsilon_i} \, \frac{1}{y_i} \, k_{i,y}(\cdot)$$

When first order conditions are satisfied, we have, moreover, $k_{i,y}(\cdot) \quad = \quad 1 - T_i(\cdot)$ and, therefore,

$$k_{i,yy}(\cdot) \quad = \quad \frac{1}{\varepsilon_i} \, \frac{1}{y_i} \left(1 - T_i(\cdot) \right) \tag{56}$$

Define the coefficients of progressivity

$$\rho^1(y) \quad = \quad -\frac{T_{11}(\cdot)}{1 - T_1(\cdot)} \, y_1 \, , \quad \rho^2(y) \quad = \quad -\frac{T_{22}(\cdot)}{1 - T_2(\cdot)} \, y_2 \, ,$$

$$\kappa^1(y) \quad = \quad \frac{k_{1,yy}(\cdot)}{k_{1,y}}\, y_1 \;, \qquad \text{and} \quad \kappa^2(y) \quad = \quad \frac{k_{2,yy}(\cdot)}{k_{2,y}}\, y_2 \;.$$

Also note that, with iso-elastic effort costs,

$$\kappa^1(y) \quad = \quad \frac{1}{\varepsilon_1} \;,$$

and

$$\kappa^2(y) \quad = \quad \frac{1}{\varepsilon_2} \;.$$

Note that, using this notation, $det(H)$, can be written as

$$det(H) \quad = \quad \frac{1-T_1(\cdot)}{y_1}\frac{1-T_2(\cdot)}{y_2}\Big(\rho^1(y)-\frac{1}{\varepsilon_1}\Big)\Big(\rho^2(y)-\frac{1}{\varepsilon_2}\Big) - T_{12}(\cdot)^2 \;. \tag{57}$$

Also, $\bar{\beta}^0_{I,1}(y)$ and $\bar{\beta}^0_{I,2}(y)$ can be written as

$$\bar{\beta}^0_{I,1}(y) = -\frac{1}{det(H)}\Big[-T_1(\cdot)\left(\rho^2(y)-\frac{1}{\varepsilon_2}\right)\frac{1-T_2(\cdot)}{y_2} - T_2(\cdot)\,T_{12}(\cdot)\Big] \tag{58}$$

and analogously,

$$\bar{\beta}^0_{I,2}(y) = -\frac{1}{det(H)}\Big[-T_2(\cdot)\left(\rho^1(y)-\frac{1}{\varepsilon_1}\right)\frac{1-T_1(\cdot)}{y_1} - T_1(\cdot)\,T_{12}(\cdot)\Big] \;. \tag{59}$$

D.2 Proof of Proposition 4

We use the conditions $\mathcal{W}_1(y_1 \mid y_2) = 0$ and $\mathcal{W}_2(y_2 \mid y_1) = 0$ to derive a candidate solution for optimal tax rates. We then verify that, under the candidate solution it is indeed the case, that $\mathcal{W}_1(y_1 \mid y_2) = 0$ and $\mathcal{W}_2(y_2 \mid y_1) = 0$.

Fix $(y_1, y_2) \in Y'$. With extensive margin elasticities of zero, and using part i) of Assumption 3, the two optimality conditions $\mathcal{W}_1(y_1 \mid y_2) = 0$ and $\mathcal{W}_2(y_2 \mid y_1) = 0$ can be written as

$$\bar{\beta}^0_{I,1}(y) f(y_1 \mid y_2) + \Big(1 - \bar{g}_1\Big)\Big(1 - F(y_1 \mid y_2)\Big) = 0 \;, \tag{60}$$

and

$$\bar{\beta}^0_{I,2}(y) f(y_2 \mid y_1) + \Big(1 - \bar{g}_2\Big)\Big(1 - F(y_2 \mid y_1)\Big) = 0 \;. \tag{61}$$

where we have been using that $f(y) = f(y_1 \mid y_2)\ f_2(y_2)$ in (60) and in (61) that $f(y) = f(y_2 \mid y_1)\ f_1(y_1)$. Using part ii) of Assumption 3 this can be rewritten as

$$\bar{\beta}^0_{I,1}(y)\ \frac{1}{y_1} = -\Big(1 - \ \bar{g}_1\Big)\frac{1}{\bar{\alpha}_1}\ , \tag{62}$$

and

$$\bar{\beta}^0_{I,2}(y)\ \frac{1}{y_2} = -\Big(1 - \ \bar{g}_2\Big)\frac{1}{\bar{\alpha}_2}\ . \tag{63}$$

where, with iso-elastic effort costs, $\bar{\beta}^0_{I,1}(y)$ and $\bar{\beta}^0_{I,2}(y)$ are characterized in Equations (57)-(59) above.

Substituting all these expressions into Equations (62) and (63) yields a system of differential equations (recall that the second derivatives of the tax function appear in all the ρ-terms.) Rather than working with that system, we hypothesize that, under Assumption 3, there is a solution that is piecewise linear. Under this hypothesis all the ρ-terms and T_{12} are zero with the implications that

$$\beta^{d0}_{I,1}(\omega) = -\frac{T_1(y)}{1 - T_1(y)}\ y_1\ \varepsilon_1\ , \tag{64}$$

and

$$\beta^{d0}_{I,2}(\omega) = -\frac{T_2(y)}{1 - T_2(y)}\ y_2\ \varepsilon_2\ . \tag{65}$$

Substituting (64) and (65) into (62) and (63) yields

$$\frac{T_1(y)}{1 - T_1(y)} \quad = \quad \Big(1 - \ \bar{g}_1\Big)\ \frac{1}{\bar{\alpha}_1}\ \frac{1}{\varepsilon_1}\ , \tag{66}$$

and

$$\frac{T_2(y)}{1 - T_2(y)} \quad = \quad \Big(1 - \ \bar{g}_2\Big)\ \frac{1}{\bar{\alpha}_2}\ \frac{1}{\varepsilon_2}\ . \tag{67}$$

and hence

$$T^0_{y_1}(y) = t^*_1 \quad \text{and} \quad T^0_{y_2}(y) = t^*_2 \,. \tag{68}$$

To complete the proof, we finally note that if we choose for all $y \in Y'$, $T^0_{y_1}(y) = t^*_1$ and $T^0_{y_2}(y) = t^*_2$, then this implies that the tax system is indeed piecewise linear: When marginal tax rates are constant, second derivatives are zero. Moreover, Equations (62) and (63) hold. Under Assumption 3 this implies that $\mathbf{W}_1(y_1 \mid y_2) = 0$ and $\mathbf{W}_2(y_2 \mid y_1) = 0$, for all $y \in Y'$, as had to be shown.

E Proof of Propositions 5 and 6

Preliminaries. Given a box $(y, l) = [y_1, y_1 + \ell_1] \times [y_2, y_2 + \ell_2]$ we introduce the following shorthands:

$$D^W_1(y, \ell) \quad = \quad \int_{y_2}^{y_2+\ell_2} \int_{y_1}^{y_1+\ell_1} \mathbf{W}_1(s_1 \mid s_2)\, ds_1\, ds_2 \,,$$

$$D^W_2(y, \ell) \quad = \quad \int_{y_2}^{y_2+\ell_2} \int_{y_1}^{y_1+\ell_1} \mathbf{W}_2(s_2 \mid s_1)\, ds_1\, ds_2 \,,$$

and

$$D^W_J(y, \ell) \quad = \quad \int_{y_2}^{y_2+\ell_2} \int_{y_1}^{y_1+\ell_1} \mathbf{W}_J(s_1, s_2)\, ds_1\, ds_2 \,.$$

We define $D^R_1(y, \ell)$, $D^R_2(y, \ell)$ and $D^R_J(y, \ell)$ analogously.

Note the equivalence of the following two statements:

A. For all (y, l), $D^W_1(y, \ell) = D^W_2(y, \ell) = D^W_J(y, \ell) = 0$.

B. For all y, $\mathbf{W}_1(y_1 \mid y_2) = \mathbf{W}_2(y_2 \mid y_1) = \mathbf{W}_J(s_1, s_2) = 0$.

Trivially, this equivalence extends to tax revenue.

Proof of Proposition 5. We first show that statement 1 implies statement 2. The proof is by contradiction. Suppose statement 2 is false. Then, by the equiv-

alence, of statements A and B, there is a box (y, ℓ) so that either $D_1^W(y,\ell) \neq 0$, $D_2^W(y,\ell) \neq 0$, or $D_{12}^W(y,\ell) \neq 0$. This implies that there is a perturbation h with $W_s(0,h) > 0$, thereby contradicting statement 1.

We now show that statement 2 implies statement 1. When statement 2 holds, then $\bar{\mathbf{W}}_1(y_1) = \bar{\mathbf{W}}_2(y_2) = 0$, for all y, since $\bar{\mathbf{W}}_1(y_1) = \int_0^{\bar{y}_2} \mathbf{W}_1(y_1 \mid y_2)dy_2$ and $\bar{\mathbf{W}}_2(y_2) = \int_0^{\bar{y}_1} \mathbf{W}_2(y_2 \mid y_1)dy_1$. By Equation (25) in Corollary 1, $\bar{\mathbf{W}}_1(y_1) = \bar{\mathbf{W}}_2(y_2) = \mathbf{W}^J(y) = 0$, for all y, imply that $W_s(0,h) = 0$, for all h.

Proof of Proposition 6. The equivalence of statements 1. and 2. follows from Proposition 5. If statement 3. is true then, for all y, $\mathbf{R}^J(0) = 0$ and also $\mathbf{R}_1(y_1 \mid y_2) = \mathbf{R}_2(y_2 \mid y_1) = 0$. Thus, statement 2. follows. It remains to be shown that statement 2 implies statement 3. Note that $\mathbf{R}^j(y)$ can be written as

$$\begin{aligned} \mathbf{R}^j(y) &= \int_{y_2}^{\bar{y}_2} \bar{\mathcal{R}}_1(y_1, s_2)ds_2 + \int_{y_1}^{\bar{y}_1} \Big(\bar{\mathcal{R}}_2(s_1, y_2)ds_1 + \int_{y_2}^{\bar{y}_2} \bar{\mathcal{R}}_b(s_1, s_2)ds_2\Big)ds_1 \\ &= \int_{y_2}^{\bar{y}_2} \bar{\mathcal{R}}_1(y_1, s_2)ds_2 + \int_{y_1}^{\bar{y}_1} \mathbf{R}_2(y_2 \mid s_1)ds_1 \ . \end{aligned} \tag{69}$$

Now, with $\mathbf{R}_2(y_2 \mid s_1) = 0$ for all y_2 and s_1, it follows that $\int_{y_2}^{\bar{y}_2} \bar{\mathcal{R}}_1(y_1, s_2)ds_2 = 0$, for all $y_1$and y_2 and hence $\bar{\mathcal{R}}_1(y) = 0$, for all y. A symmetric argument can be used to show that $\bar{\mathcal{R}}_2(y) = 0$, for all y. But then $\mathbf{R}^j(y) = 0$, for all y requires that $\int_{y_2}^{\bar{y}_2} \bar{\mathcal{R}}_b(s_1, s_2)ds_2 = 0$ for all s_1 and y_2. This can only be true if $\bar{\mathcal{R}}_b(y) = 0$, for all y.

Supplementary Material

F Completing the Proof of Proposition 1

Case 3: Suppose that the tax function jumps upwards at y_1', and that the marginal tax rate also jumps upwards. Formally, $\lim_{\epsilon\to 0^+} T(y_1'+\epsilon, y_2) > \lim_{\epsilon\to 0^-} T(y_1'+\epsilon, y_2)$ and $\lim_{\epsilon\to 0^+} T_{y_1}(y_1'+\epsilon, y_2) > \lim_{\epsilon\to 0^-} T_{y_1}(y_1'+\epsilon, y_2)$.

Suppose there is a (possibly empty) subset of types $\Omega_1^B(y_1') \subset \Omega_1$ who choose y_1'. Again, Equation (36) has to hold. Otherwise, individuals were better off lowering y_1 slightly below y_1'. In addition, choosing y_1' must be better than choosing some $y_1 > y_1'$. Formally, this requires that

$$u(c') - k_1(y_1', \omega_1) - v_{\bar{n}}(\omega_1) \quad \geq \quad 0\,, \tag{70}$$

where, for given y_2, $v_{\bar{n}}(\omega_1) = \max_{y_1 > y_1'} u(b + y_1 + y_2 - T(y_1, y_2)) - k(y_1, \omega_1)$ is the value function associated with the problem of choosing y_1 optimally after the notch. It follows from an application of the envelope theorem and the single-crossing condition that the left hand side of (70) is decreasing in ω_1: To see this, let $Y_{\bar{n}}(\omega_1)$ be the (not necessarily single-valued) set of optimal choices subject to the constraint $y_1 > y_1'$. Let $\underline{y}_1^*$ be the smallest element of this set. It follows from Corollary 4 in Milgrom and Segal (2002), that the derivative of left hand side of (70) is bounded from above by

$$k_{\omega_1}(\underline{y}_1^*, \omega) - k_{\omega_1}(y_1', \omega_1) \quad = \quad \int_{y_1'}^{\underline{y}_1^*} k_{y_1\omega_1}(s, \omega_1)\, ds\,.$$

This expression is negative since the cross derivative of the effort cost function is negative by assumption.

Let $\underline{\omega}_1$ be the level of ω_1 for which (36) holds as an equality. Let $\bar{\omega}_1$ be the level of ω_1 for which (70) holds as an equality. The assumptions that define Case 3 imply that $\underline{\omega}_1 < \bar{\omega}_1$, see Figure 2. Again, suppose that $\Omega_1^B(y_1')$ contains types (ω_1, ω_2) so that $\omega_1 \in]\underline{\omega}_1, \bar{\omega}_1[$. For such a type, conditions (36) and (70) hold as strict inequalities, and as the Figure shows, y_1' is the unique solution to the problem of maximizing utility given y_2.

It is also a unique solution for type $\underline{\omega}_1$, but the conditions of strict bunching does not hold for this type. Strict bunching is satisfied, by contrast, for type $\bar{\omega}_1$. This type, however, has a second optimal choice. Thus, there is at most one type in $\Omega_1^B(y_2)$ with multiple solutions and at most one type violating strict bunching.

Case 4: Suppose that the tax function jumps downwards at y_1', and that the marginal tax rate jumps upwards. Formally, $\lim_{\epsilon\to 0^+} T(y_1'+\epsilon, y_2) < \lim_{\epsilon\to 0^-} T(y_1'+\epsilon, y_2)$ and $\lim_{\epsilon\to 0^+} T_{y_1}(y_1'+\epsilon, y_2) > \lim_{\epsilon\to 0^-} T_{y_1}(y_1'+\epsilon, y_2)$.

Suppose there is a (possibly empty) subset of types $\Omega_d^B(y_1) \subset \Omega_1$ who choose y_1'. Again, Equation (35) has to hold. Otherwise, individuals were better off raising y_1 slightly above y_1'. In addition, choosing y_1' must be better than choosing some $y_1 < y_1'$. Formally, this requires that

$$u(c') - k_1(y_1', \omega_1) \; - \; v_{\underline{n}}(\omega_1) \quad \geq \quad 0 \; , \tag{71}$$

where, for given y_2, $v_{\underline{n}}(\omega_1) = \max_{y_1<y_1'} \; u(b + y_1 + y_2 - T(y_1, y_2)) - k(y_1, \omega_1)$ is the value function associated with the problem of choosing y_1 optimally before the notch, and $Y_{\underline{n}}(\omega_1)$ is the (not necessarily single-valued) set of maximizers. It follows from an application of the envelope theorem and the single-crossing condition that the left hand side of (71) is decreasing in ω_1: Let $\bar{y}_1^*$ be the largest element of $Y_{\underline{n}}(\omega_1)$. It follows from Corollary 4 in Milgrom and Segal (2002), that the derivative of the left hand side of (71) is bounded from below by

$$-\Big(k_{\omega_1}(y_1', \omega_1) - k_{\omega_1}(\bar{y}_1^*, \omega)\Big) \quad = \quad -\int_{\bar{y}_1^*}^{y_1'} k_{y_1\omega_1}(s, \omega_1)\, ds \; .$$

This expression is positive since the cross derivative of the effort cost function is negative.

Let $\bar{\omega}_1$ be the level of ω_1 for which (36) holds as an equality. Let $\underline{\omega}_1$ be the level of ω_1 for which (71) holds as an equality. The assumptions that define Case 4 imply that $\underline{\omega}_1 < \bar{\omega}_1$, see Figure 2.

Now suppose that $\Omega_d^B(y_1')$ contains $\omega_1 \in]\underline{\omega}_1, \bar{\omega}_1[$. For such a type, conditions (35) and (36) hold as strict inequalities, with the implication that the condition of strict bunching is satisfied. Moreover, as the figure shows, y_1' is the unique solution to the problem of maximizing utility given y_2. Thus, there are at most two types

in $\Omega_1^B(y_2)$ with multiple solutions or with a violation of strict bunching.

Case 5: Suppose that the tax function jumps upwards at y_1', and that the marginal tax rate jumps downwards, see Figure 2. Alternatively, suppose that the tax function jumps downwards at y_1', and that the marginal tax rate also jumps downwards. This case is analogous to Case 2 in that it creates a local convexity in the budget set, with the implication that choosing y_1' is dominated. Thus, $\Omega_d^B(y_1') = \emptyset$.

G Supplement to the Proof of Proposition 2

We seek a characterization of

$$\Delta_x^0(y^{jk}, \ell_1, \ell_2) \quad := \quad \mathbf{E}_\omega[r_x^0(\omega)\ \mathbf{1}(\omega \in \Omega^{jk}(0,h))]\ ,$$

where $\Omega^{jk}(0,h) = \{\omega \mid y_d^0(\omega) \in B^{jk}\}$.

G.1 Characterizing $\Omega^{jk}(0,h)$

Assume, henceforth, that $\Omega^{jk}(0,h)$ is non-empty. Let $B_1^{jk} = [y_1^{jk}, y_1^{jk} + \ell_1]$, and

$$\underline{\omega}_1(y_1^{jk}) \quad := \quad \inf\{\omega_1 \in \Omega_1 \mid \exists\, \omega_2 \text{ s.t. } y_{d1}^0(\omega_1, \omega_2) \geq y_1^{jk}\}\ .$$

Analogously, define

$$\overline{\omega}_1(y_1^{jk} + \ell_1) \quad := \quad \sup\{\omega_1 \in \Omega_1 \mid \exists\, \omega_2 \text{ s.t. } y_{d1}^0(\omega_1, \omega_2) \leq y_1^{jk} + \ell_1\}\ .$$

Consider some fixed $\omega_1 \in \Omega_1(B_1^{jk}) = [\underline{\omega}_1(y_1^{jk}), \overline{\omega}_1(y_1^{jk} + \ell_1)]$. There is then a non-empty subset $\Omega_2'[B_1^{jk}, \omega_1]$ of Ω_2 so that $\omega_2 \in \Omega_2'[B_1^{jk}, \omega_1]$ implies $y_{d1}^0(\omega_1, \omega_2) \in [y_1^{jk}, y_1^{jk} + \ell_1]$. There is then also a subset $\Omega_2[B^{jk}, \omega_1]$ of $\Omega_2'[B_1^{jk}, \omega_1]$ so that $\omega_2 \in \Omega_2[B^{jk}, \omega_1]$ implies $y_{d2}^0(\omega_1, \omega_2) \in [y_2^{jk}, y_2^{jk} + \ell_2]$. Denote, respectively, by $\underline{\omega}_2(y_2^{jk} \mid \omega_1)$ and $\overline{\omega}_2(y_2^{jk} + \ell_2 \mid \omega_1)$ the inf and sup of this set.

G.2 Characterizing $\beta_{I,x}^0(\omega)$, $\beta_{E,x}^0(\omega)$ and $\mu_x^0(\omega)$

Fix an interior box B^{jk}. We seek to characterize $r_x^0(\omega) = r_x(0, h, \omega)$, for $x \in \{sb^{jk}, s\tau_1^{jk}, s\tau_2^{jk}\}$. To get there we will first characterize $r_x(s, h, \omega)$ for arbitrary s

and then evaluate for $s = 0$. We proceed as follows: We look at the perturbation

$$h^{jk}(y) = b^{jk} + \tau_1^{jk}(y_1 - y_1^{jk}) + \tau_2^{jk}(y_2 - y_2^{jk})$$

that applies when $y \in B^{jk}$. Recall that, conditional on $y \in B^{jk}$, the perturbed tax payment is then given by

$$T(y) + s\ h^{jk}(y)\ .$$

We define $r_s(s, h^{jk}, \omega)$ as the change in revenue due to individuals with type ω when the scalar s is changed at the margin. We use the shorthand $r^0(h^{jk}, \omega) := r_s(0, h^{jk}, \omega)$. We show that $r^0(h^{jk}, \omega)$ can be written as

$$r^0(\omega, h^{jk}) = b^{jk}\ r_b^0(\omega) + \tau_1^{jk}\ r_1^0(\omega) + \tau_2^{jk}\ r_2^0(\omega)$$

and provide a characterization of $r_b^0(\omega)$, $r_1^0(\omega)$, and $r_2^0(\omega)$.

From Equation (9) we have

$$\begin{aligned} r(s, h^{jk}, \omega) \ &= \ \sigma^{12}(s, h^{jk}, \omega)\ \Big(T(y_d^*(s, h^{jk}, \omega)) + s\ h^{jk}(y_d^*(s, h^{jk}, \omega))\Big) \\ &\quad +\sigma^1(s, h^{jk}, \omega)\ \Big(T(y_{1s}^*(s, h^{j'0}, \omega), 0) + s\ h^{j'0}(y_{1s}^*(s, h, \omega), 0)\Big) \\ &\quad +\sigma^2(s, h^{jk}, \omega)\ \Big(T(0, y_{2s}^*(s, h^{0,k'}, \omega)) + s\ h^{0,k'}(0, y_{2s}^*(s, h, \omega))\Big)\ , \end{aligned} \tag{72}$$

where $j'0$ and $0k'$ index the non-interior boxes that contain the earnings choices of those in the ω-population which have fixed costs φ so that these agents either choose to have earnings only from activity 1 or to have earnings only from activity 2. We are interested in the ceteris paribus effects that a change in s has on h^{jk}, which amounts to assuming that $h^{j'0} = h^{0k'} \equiv 0$. This also implies that choices in those non-interior boxes are not affected by the perturbation that we consider. Thus we simply let $y_{1s}^*(s, h, \omega) = y_{1s}^0(\omega_1)$ and $y_{2s}^*(s, h, \omega) = y_{2s}^0(\omega_2)$ and employ

$$\begin{aligned} r(s, h^{jk}, \omega) \ &= \ \sigma^{12}(s, h^{jk}, \omega)\ \Big(T(y_d^*(s, h^{jk}, \omega)) + s\ h^{jk}(y_d^*(s, h^{jk}, \omega))\Big) \\ &\quad +\sigma^1(s, h^{jk}, \omega)\ T(y_{1s}^0(\omega_1), 0)\ +\ \sigma^2(s, h^{jk}, \omega)\ T(0, y_{2s}^0(\omega_2))\ . \end{aligned} \tag{73}$$

The choices of types $\omega \in \Omega'$ with $y_d^*(\cdot) \in B^{jk}$ either satisfy strict bunching or first order conditions. By Proposition 1, types in Ω' have unique optima, so that the specification of the tax system in other boxes does not affect the choice of type ω. Thus, we can write

$$\begin{aligned} y_d^*(s, h^{jk}, \omega) &= y_d^*(sb^{jk}, s\tau_1^{jk}, s\tau_2^{jk}, \omega) \\ &= \Big(y_{d1}^*(sb^{jk}, s\tau_1^{jk}, s\tau_2^{jk}, \omega), y_{d2}^*(sb^{jk}, s\tau_1^{jk}, s\tau_2^{jk}, \omega)\Big). \end{aligned} \tag{74}$$

For later reference, we denote the derivative of y_{di}^* with respect to its first, second and third argument, respectively, by $y_{di,b}^*(\cdot)$, $y_{di,1}^*(\cdot)$ and $y_{di,2}^*(\cdot)$.[58] If we evaluate these expressions at the status quo, i.e. for $s = 0$, we write $y_{di,b}^0(\omega)$, $y_{di,1}^0(\omega)$ and $y_{di,2}^0(\omega)$.

For a general reform direction h, the fraction of type ω-agents choosing earnings from both activities, only one activity, or no earnings at all can be written as functions of payoff differences. E.g, $\sigma^{12}(s, h, \omega)$ can be written as a function of $v^{12}(s, h, \omega) - v^1(s, h, \omega_1)$, $v^{12}(s, h, \omega) - v^2(s, h, \omega_2)$, and $v^{12}(s, h, \omega) - v^0(\delta)$. Likewise, $\sigma^1(s, h, \omega)$ can be written as a function of $v^1(s, h, \omega_1) - v^{12}(s, h, \omega)$, $v^1(s, h, \omega_1) - v^2(s, h, \omega_2)$, and $v^1(s, h, \omega_1) - v^0(\delta)$, etc. Henceforth, for the purpose of the given ceteris paribus exercise, we treat $v^1(\cdot)$, $v^2(\cdot)$ and $v^0(\cdot)$ as constants and simply write v^1, v^2 and v^0. For later reference, note that, by the envelope theorem,

$$\begin{aligned} v_s^{12}(0, h^{jk}) &= -u(c_d^0(\omega))h^{jk}(y_d^0(\omega)) \\ &= -u(c_d^0(\omega))\Big(b^{jk} + \tau_1^{jk}\big(y_{d1}^0(\omega) - y_1^{jk}\big) + \tau_2^{jk}\big(y_{d1}^0(\omega) - y_1^{jk}\big)\Big), \end{aligned} \tag{75}$$

for $c_d^0(\omega) := b + y_{d1}^0(\omega) + y_{d2}^0(\omega) - T(y_d^0(\omega))$.

With v^1, v^2 and v^0 held fixed, σ^{12} depends only on v^{12}. Thus, we can write

$$\sigma^{12}(s, h^{jk}, \omega) = \sigma^{12}\Big(v^{12}(s, h^{jk}) - v^1, v^{12}(s, h^{jk}) - v^2, v^{12}(s, h^{jk}) - v^0\Big) \tag{76}$$

We denote the derivatives of σ^{12} with respect to the first, second and third argument

58. If there is strict bunching in both activities these derivatives are all zero. If both choices satisfy first order conditions, these derivatives are all different from zero. If the satisfy strict bunching for one activity and a first order condition for the other some of these derivatives are zero and others are different from zero.

by $\sigma_1^{12}(\cdot)$, $\sigma_2^{12}(\cdot)$ and $\sigma_0^{12}(\cdot)$, respectively. If we evaluate these expressions at the status quo, i.e. at $s = 0$, we write $\sigma_1^{12,0}(\omega)$, $\sigma_2^{12,0}(\omega)$ and $\sigma_0^{12,0}(\omega)$. We write $\sigma_t^{12}(\cdot)$ and $\sigma_t^{12,0}(\omega)$ for the total differential that captures the effect of marginal changes in $v^{12}(\cdot)$ on the fraction of type ω-individuals with earnings from both activities. Thus,

$$\sigma_t^{12}(\cdot) \quad = \quad \sigma_1^{12}(\cdot) + \sigma_2^{12}(\cdot) + \sigma_0^{12}(\cdot) \; ,$$

and

$$\sigma_t^{12,0}(\omega) \quad = \quad \sigma_1^{12,0}(\omega) + \sigma_2^{12,0}(\omega) + \sigma_0^{12,0}(\omega) \; .$$

We use analogous notation, when dealing with

$$\sigma^1(s, h, \omega) = \sigma^1\Big(v^1 - v^{12}(s, h^{jk}), v^1 - v^2, v^1 - v^0\Big) \tag{77}$$

and

$$\sigma^2(s, h, \omega) = \sigma^2\Big(v^2 - v^{12}(s, h^{jk}), v^2 - v^1, v^2 - v^0\Big) \tag{78}$$

We can now rewrite Equation (73) as

$$\begin{aligned} r(s, h^{jk}, \omega) \quad &= \quad \sigma^{12}(\cdot) \left(T(y_d^*(\cdot)) + s \; h^{jk}(y_d^*(\cdot))\right) \\ &\quad + \sigma^1(\cdot) \; T(y_{1s}^0(\omega_1), 0) \; + \; \sigma^2(\cdot) \; T(0, y_{2s}^0(\omega_2)) \; . \end{aligned} \tag{79}$$

and compute

$$r_s(0, h^{jk}, \omega) \quad = \quad \mathbf{b}_I + \mathbf{m} + \mathbf{b}_E \; , \tag{80}$$

where

$$\begin{aligned} \mathbf{b}_I \quad = \quad \sigma^{12,0}(\omega) \Big[&T_1(y_d^0(\omega)) \left(b^{jk} \; y_{d1,b}^0(\omega) + \tau_1^{jk} \; y_{d1,1}^0(\omega) + \tau_2^{jk} \; y_{d1,2}^0(\omega)\right) \\ &+ T_2(y_d^0(\omega)) \left(b^{jk} \; y_{d2,b}^0(\omega) + \tau_1^{jk} \; y_{d2,1}^0(\omega) + \tau_2^{jk} \; y_{d1,2}^0(\omega)\right)\Big] \end{aligned} \tag{81}$$

captures the effect of intensive margin behavioral responses on tax revenue. Note

that we can write

$$\mathbf{b}_I = b^{jk}\,\beta^0_{I,b}(\omega) + \tau_1^{jk}\,\beta^0_{I,1}(\omega) + \tau_2^{jk}\,\beta^0_{I,2}(\omega) \tag{82}$$

where

$$\beta^0_{I,b}(\omega) = \sigma^{12,0}(\omega)\left[T_1(y_d^0(\omega))\; y^0_{d1,b}(\omega) + T_2(y_d^0(\omega))y^0_{d2,b}(\omega)\right],$$

$$\beta^0_{I,1}(\omega) = \sigma^{12,0}(\omega)\left[T_1(y_d^0(\omega))\; y^0_{d1,1}(\omega) + T_2(y_d^0(\omega))y^0_{d2,1}(\omega)\right],$$

and

$$\beta^0_{I,2}(\omega) = \sigma^{12,0}(\omega)\left[T_1(y_d^0(\omega))\; y^0_{d1,2}(\omega) + T_2(y_d^0(\omega))y^0_{d2,2}(\omega)\right].$$

The mechanical effect on tax revenue is given by

$$\begin{aligned}\mathbf{m} &= \sigma^{12,0}(\omega)\; h^{jk}(y_d^0(\omega)) \\ &= \sigma^{12,0}(\omega)\Big(b^{jk} + \tau_1^{jk}(y^0_{d1}(\omega) - y_1^{jk}) + \tau_2^{jk}(y^0_{d2}(\omega) - y_2^{jk})\Big)\end{aligned} \tag{83}$$

and can be written as

$$\mathbf{m} = b^{jk}\,\mu_b^0(\omega) + \tau_1^{jk}\,\mu_1^0(\omega) + \tau_2^{jk}\,\mu_2^0(\omega), \tag{84}$$

where

$$\mu_b^0(\omega) = \sigma^{12,0}(\omega)\,,$$

$$\mu_1^0(\omega) = \sigma^{12,0}(\omega)\big(y^0_{d1}(\omega) - y_1^{jk}\big),$$

and

$$\mu_2^0(\omega) = \sigma^{12,0}(\omega)\big(y^0_{d2}(\omega) - y_2^{jk}\big)\,.$$

Finally, the effect of extensive margin behavioral responses is given by

$$\begin{aligned}\mathbf{b}_E = -u'(c_d^0(\omega))h^{jk}(y_d^0(\omega))\,\big[&\sigma_t^{12,0}(\omega)\; T(y_d^0(\omega)) - \sigma_{12}^{1,0}(\omega)\; T(y^0_{1s}(\omega_1), 0) \\ &-\sigma_{12}^{2,0}(\omega)\; T(0, y^0_{2s}(\omega_2))\big]\end{aligned} \tag{85}$$

Note that we can write

$$\mathbf{b}_E \;=\; b^{jk}\;\beta^0_{E,b}(\omega) + \tau_1^{jk}\;\beta^0_{E,1}(\omega) + \tau_2^{jk}\;\beta^0_{E,2}(\omega) \tag{86}$$

where

$$\begin{aligned}\beta^0_{E,b}(\omega) \;=\; -u'(c^0_d(\omega))\Big[&\sigma_t^{12,0}(\omega)\;T(y^0_d(\omega)) - \sigma_{12}^{1,0}(\omega)\;T(y^0_{1s}(\omega_1),0)\\ &-\sigma_{12}^{2,0}(\omega)\;T(0,y^0_{2s}(\omega_2)),\Big]\end{aligned} \tag{87}$$

$$\begin{aligned}\beta^0_{E,1}(\omega) \;=\; -u'(c^0_d(\omega))\big(y^0_{d1}(\omega) - y_1^{jk}\big)\Big[&\sigma_t^{12,0}(\omega)\;T(y^0_d(\omega)) - \sigma_{12}^{1,0}(\omega)\;T(y^0_{1s}(\omega_1),0)\\ &-\sigma_{12}^{2,0}(\omega)\;T(0,y^0_{2s}(\omega_2)),\Big]\end{aligned} \tag{88}$$

and

$$\begin{aligned}\beta^0_{E,2}(\omega) \;=\; -u'(c^0_d(\omega))\big(y^0_{d2}(\omega) - y_2^{jk}\big)\Big[&\sigma_t^{12,0}(\omega)\;T(y^0_d(\omega)) - \sigma_{12}^{1,0}(\omega)\;T(y^0_{1s}(\omega_1),0)\\ &-\sigma_{12}^{2,0}(\omega)\;T(0,y^0_{2s}(\omega_2)).\Big]\end{aligned} \tag{89}$$

Upon collecting terms we obtain the following characterization

$$r_s(0,h^{jk},\omega) \;=\; b^{jk}\;r^0_b(\omega) + \tau_1^{jk}\;r^0_1(\omega) + \tau_2^{jk}\;r^0_2(\omega) \tag{90}$$

where, for any $x \in \{b,1,2\}$,

$$r^0_x(\omega) = \beta^0_{I,x}(\omega) + \mu^0_x(\omega) + \beta^0_{E,x}(\omega)\;. \tag{91}$$

H Supplement to the Proof of Proposition 3

H.1 Preliminaries

H.1.1 Further implications of the analysis in Section G.2

Implications of $\ell \to 0$. Note that $0 \le y^0_{d2}(\omega) - y_2^{jk} \le \ell_2$ and $0 \le y^0_{d1}(\omega) - y_1^{jk} \le \ell_1$. Hence, $\ell \to 0$ implies $\mu^0_1(\omega) \to 0$, $\mu^0_2(\omega) \to 0$, $\beta^0_{E,1}(\omega) \to 0$ and $\beta^0_{E,2}(\omega) \to 0$. Consequently,

$$r^0_1(\omega) \;\to\; \beta^0_{I,1}(\omega) \quad \textit{and} \quad r^0_2(\omega) \;\to\; \beta^0_{I,2}(\omega)\;. \tag{92}$$

Implications of bunching at a mass point. If y is a mass point, then there is a positive mass of agents with $\omega \in \Omega'$ so that they satisfy the strict bunching condition both in their choice of y_1 and in their choice of y_2. Consequently, they show no behavioral responses to marginal changes of the tax burden or of marginal tax rates. Therefore,

$$r_1^0(\omega) \;=\; 0 \quad \textit{and} \quad r_2^0(\omega) \;=\; 0\,. \tag{93}$$

H.1.2 Types and incomes associated with a box

Types in the box. Who chooses earnings $y_d^0(\omega)$ in B^{jk}? By Proposition 1, there is a largest value $\overline{\omega}_1(y_1^{jk} + \ell_1)$ and a smallest value $\underline{\omega}_1(y_1^{jk})$ of ω_1 consistent with $y_d^0(\omega) \in B^{jk}$. Given ω_1 in this range, there is a largest value $\overline{\omega}_2(y_2^{jk} + \ell_2, \omega_1)$ and a smallest value $\underline{\omega}_2(y_2^{jk}, \omega_1)$ of ω_2 consistent with $y_d^0(\omega) \in B^{jk}$.

Notation. We denote by Y^M be the set of all mass points. We denote by $f^M(y_m)$ the mass of taxpayers choosing $y = y_m$. We refer to the income density that results from the atomless part of the distribution as f^A. This density captures how the mass of ω-types in a "very narrow" box with y as the lower left corner is affected by marginal changes of ℓ_1 and ℓ_2.[59]

From types to incomes. For $y_m \in Y^M$, $\mathcal{R}_x^M(y_m) := f^M(y_m)\; \bar{r}_x^0(y_m)$ gives the revenue implications of the behavioral responses of all those types ω who bunch at y_m . The interpretation is that if a parameter x of the tax function changes at y_m, then the revenue effect is given by $\mathcal{R}_x^M(y_m)$. We define an analogous expression for incomes in the atomless part of $\mathcal{Y}$: For $y \notin Y^M$,

$$\mathcal{R}_x^A(y) = f^A(y)\; \bar{r}_x^0(y) \quad \text{with} \quad \bar{r}_x^0(y) := r_x^0\left(\bar{\omega}(y)\right)$$

where $\bar{\omega}(y) = (\overline{\omega}_1(y_1), \overline{\omega}_2(y_2 \mid \overline{\omega}_1(y_1))$ is "the highest" type choosing y. The decomposition in (91) implies that, for any y, we can write

$$\bar{r}_x^0(y) \quad = \quad \bar{\beta}_{I,x}^0(y) + \bar{\beta}_{E,x}^0(y) + \bar{\mu}_x^0(y).$$

59. Equation (94) below makes explicit how f^A is derived from $\tilde{F}$.

The analysis in Appendix G.2 moreover implies that $\bar{\beta}_{I,x}(y) = 0$ if y is a mass point, and that $\bar{\beta}_{E,1}(y)$, $\bar{\mu}_1(y)$, $\bar{\beta}_{E,2}(y)$, and $\bar{\mu}_2(y)$ all vanish as $\ell \to 0$. The following Lemma summarizes this discussion.

H.1.3 Implications of making the grid finer and finer

Lemma 2 *Consider a sequence of grids indexed by n. Let $\ell_1^n = \frac{1}{2}\ell_1^{n-1}$ and $\ell_2^n = \frac{1}{2}\ell_2^{n-1}$. Index the functions $\mathcal{R}_x$ with a superscript n to emphasize their dependence on a given grid. For $j \in \{A, M\}$, denote $\bar{\mathcal{R}}_x^j(y) := \lim_{n\to\infty} \mathcal{R}_x^{nj}(y)$. Then, if y is a mass point, $\bar{\mathcal{R}}_1^M(y) = \bar{\mathcal{R}}_2^M(y) = 0$ and $\bar{\mathcal{R}}_b^M(y) = f^M(y)\Big(\bar{\beta}_{E,b}^0(y) + \bar{\mu}_b^0(y)\Big)$. If y is not a mass point, then $\bar{\mathcal{R}}_1^A(y) = f^A(y)\,\bar{\beta}_{I,1}^0(y)$, $\bar{\mathcal{R}}_2^A(y) = f^A(y)\,\bar{\beta}_{I,2}^0(y)$ and $\bar{\mathcal{R}}_b^A(y) = f^A(y)\left(\bar{\beta}_{I,b}^0(y) + \bar{\beta}_{E,b}^0(y) + \bar{\mu}_b^0(y)\right)$.*

We define $\bar{\mathcal{R}}_b(y) := \bar{\mathcal{R}}_b^A(y) + \mathbf{1}(y \in Y_M)\bar{\mathcal{R}}_b^M(y)$, and analogously for $\bar{\mathcal{R}}_1(y)$ and $\bar{\mathcal{R}}_2(y)$. This allows us to write expressions that give the total effect from an aggregation over the income space – integrating over the atomless part and summing over the mass points – more compactly; e.g. we write

$$\int_{\mathcal{Y}} h(y)\,\bar{\mathcal{R}}_b(y)dy \;\; := \;\; \int_{\mathcal{Y}} h(y)\,\bar{\mathcal{R}}_b^A(y)dy + \int_{\mathcal{Y}} \mathbf{1}(y \in Y^M)\,h(y)\,\bar{\mathcal{R}}_b^M(y)dy$$

rather than $\int_{\mathcal{Y}} h(y)\,\bar{\mathcal{R}}_b^A(y)dy + \sum_{y_m \in Y_M} h(y_m)\bar{\mathcal{R}}_b^M(y_m)$.

H.2 Implications of $\ell \to 0$ for $\Delta_x^{0\nu}(y^{jk}, \ell_1, \ell_2)$

We turn to the characterization of

$$\Delta_x^{0\nu}(y^{jk}, \ell_1, \ell_2) \;\; = \;\; \int_{\underline{\omega}_1(y_1^{jk})}^{\overline{\omega}_1(y_1^{jk}+\ell_1)} \int_{\underline{\omega}_2(y_2^{jk}|\omega_1)}^{\overline{\omega}_2(y_2^{jk}+\ell_2|\omega_1)} r_x^0(\omega)\; d\tilde{F}^A(\omega)$$

for ℓ close to zero, where F^A is the atomless cdf over types who do not bunch. We denote by $(y_1, y_2) \mapsto f^A(y_1, y_2)$ the density that represents the joint distribution of earnings that is generated by the choices of those types who are not at mass points. The marginal densities are denoted by f_1^A and f_2^A and the marginal *cdf*s by F_1^A and F_2^A. We write $f_2^A(\cdot \mid y_1)$ and $F_2^A(\cdot \mid y_1)$ for conditional densities and *cdf*s.

The assumption that there are no mass points implies that

$$\bar{\nu}(y^{jk}) \;:=\; \int_{\underline{\omega}_1(y_1^{jk})}^{\overline{\omega}_1(y_1^{jk})} \int_{\underline{\omega}_2(y_2^{jk}|\omega_1)}^{\overline{\omega}_2(y_2^{jk}|\omega_1)} d\tilde{F}^A(\omega) \;=\; 0 \,.$$

If this expression was strictly positive there would be a mass point at y^{jk}. Note that we also have

$$\bar{\nu}(y^{jk}) \;:=\; \bar{\nu}_1(y_1^{jk})\; \bar{\nu}(y_2^{jk} \mid y_1^{jk}) \,,$$

where

$$\bar{\nu}_1(y_1^{jk}) \;:=\; \tilde{F}_1^A(\overline{\omega}_1(y_1^{jk})) - \tilde{F}_1^A(\underline{\omega}_1(y_1^{jk}))$$

and

$$\bar{\nu}(y_2^{jk} \mid y_1^{jk}) \;:=\; \mathbf{E}_{\omega_2}^A \left[\tilde{F}_2^A(\overline{\omega}_2(y_2^{jk} \mid \omega_1)) - \tilde{F}_2^A(\underline{\omega}_2(y_2^{jk} \mid \omega_1)) \mid \omega_1 \in [\underline{\omega}_1(y_1^{jk}), \overline{\omega}_1(y_1^{jk})] \right] .$$

Also note that $\bar{\nu}(y^{jk}) = 0$ implies $\bar{\nu}_1(y_1^{jk}) = 0$ or $\bar{\nu}(y_2^{jk} \mid y_1^{jk}) = 0$. In the following we proceed under the assumption that $\bar{\nu}_1(y^{jk}) = 0$. An analogous derivation could be made under the alternative assumption that $\bar{\nu}(y_2^{jk} \mid y_1^{jk}) = 0$.

Two consecutive first order Taylor approximations imply that

$$\Delta_x^{0\nu}(y^{jk}, \ell_1, \ell_2) \;\simeq\; \ell_1\, \Delta_{x,\ell_1}^{0\nu}(y^{jk}, 0, 0) + \ell_2\, \Delta_{x,\ell_2}^{0\nu}(y^{jk}, 0, 0) + \ell_2\ell_1\, \Delta_{x,\ell_2\ell_1}^{0\nu}(y^{jk}, 0, 0) \,,$$

where we have used already that, since there are no mass points,

$$\Delta_x^{0,\nu}(y^{jk}, 0, 0) \;=\; \bar{r}_x^{0,\nu}(y^{jk})\; \bar{\nu}(y^{jk}) \;=\; 0 \,,$$

for

$$\bar{r}_x^{0,\nu}(y^{jk}) \;:=\; \mathbf{E}_\omega^A[r_x^0(\omega) \mid y_d^0(\omega) = y^{jk}] \,.$$

We will see in the derivations below that also $\Delta_{x,\ell_1}^{0\nu}(y^{jk}, 0, 0) = \Delta_{x,\ell_2}^{0\nu}(y^{jk}, 0, 0) = 0$, so that, ultimately,

$$\Delta_x^{0\nu}(y^{jk}, \ell_1, \ell_2) \;\simeq\; \ell_2\ell_1\, \Delta_{x,\ell_2\ell_1}^{0\nu}(y^{jk}, 0, 0) \,,$$

Implications of $\ell_2 \to 0$. For any y_2', and any ω_1, let

$$F_2^A(y_2' \mid \omega_1) \quad := \quad \tilde{F}_2^A(\bar{\omega}_2(y_2' \mid \omega_1) \mid \omega_1) \ ,$$

where $\bar{\omega}_2(y_2' \mid \omega_1)$ is the largest value of ω_2 so that $y_{d2}^0(\omega_1, \omega_2) \leq y_2'$. Upon differentiating with respect to y_2', we obtain

$$f_2^A(y_2' \mid \omega_1) \quad := \quad \tilde{f}_2^A(\bar{\omega}_2(y_2' \mid \omega_1) \mid \omega_1) \frac{\partial}{\partial y_2'} \bar{\omega}_2(y_2' \mid \omega_1) \ .$$

Using this notation and upon differentiating

$$\Delta_x^{0,\nu}(y^{jk}, \ell_1, \ell_2) \quad = \quad \int\limits_{\underline{\omega}_1(y_1^{jk})}^{\overline{\omega}_1(y_1^{jk}+\ell_1)} \int\limits_{\underline{\omega}_2(y_2^{jk}, \omega_1)}^{\overline{\omega}_2(y_2^{jk}+\ell_2, \omega_1)} r_x^0(\omega) \, d\tilde{F}^A(\omega) \ ,$$

where $d\tilde{F}^A(\omega)$ is a shorthand for $\tilde{f}_{\omega_2}^A(\omega_2 \mid \omega_1) d\omega_2 \ \tilde{f}_{\omega_1}^A(\omega_1) d\omega_1$, with respect to ℓ_2 and evaluating the resulting expressions at $\ell_2 = 0$, we obtain

$$\Delta_{x,\ell_2}^{0,\nu}(y^{jk}, \ell_1, 0) \quad = \quad \int\limits_{\underline{\omega}_1(y_1^{jk})}^{\overline{\omega}_1(y_1^{jk}+\ell_1)} r_x^0 \left(\omega_1, \bar{\omega}_2(y_2^{jk} \mid \omega_1)\right) \ f_2^A(y_2^{jk} \mid \omega_1) \ \tilde{f}_{\omega_1}^A(\omega_1) d\omega_1 \ ,$$

For $\ell_1 = 0$ this expression can be written as

$$\Delta_{x,\ell_2}^{0,\nu}(y^{jk}, 0, 0) \quad = \quad \bar{\nu}_1(y_1^{jk}) \ \bar{r}_x^0(y_2^{jk} \mid y_1^{jk})$$

where

$$\bar{r}_x^0(y_2^{jk} \mid y_1^{jk}) \quad := \quad \mathbf{E}_{\omega_1}^A \left[r_x^0 \left(\omega_1, \bar{\omega}_2(y_2^{jk} \mid \omega_1)\right) \ f_{y_2}^A(y_2^{jk} \mid \omega_1) \mid \omega_1 \in [\underline{\omega}_1(y_1^{jk}), \overline{\omega}_1(y_1^{jk})] \right] \ .$$

Hence, $\bar{\nu}_1(y_1^{jk}) = 0$ implies that $\Delta_{x,\ell_2}^{0\nu}(y^{jk}, 0, 0) = 0$. A related argument can be used to show that , $\bar{\nu}(y_1^{jk}) = 0$ implies that $\Delta_{x,\ell_1}^{0\nu}(y^{jk}, 0, 0) = 0$. To see this, note that

$$\Delta_{x,\ell_1}^{0\nu}(y^{jk}, 0, 0) \quad = \quad \left(\int\limits_{\underline{\omega}_2(y_2^{jk}, \bar{\omega}_1(y_1^{jk}))}^{\overline{\omega}_2(y_2^{jk}, \bar{\omega}_1(y_1^{jk}))} r_x^0(\omega) \tilde{f}_{\omega_2}^A(\omega_2 \mid \bar{\omega}_1(y_1^{jk})) d\omega_2 \right) \tilde{f}_{\omega_1}^A(\bar{\omega}_1(y_1^{jk})) \frac{\partial}{\partial y_1^{jk}} \bar{\omega}_1(y_1^{jk}) \ .$$

Since

$$\bar{\nu}(y^{jk}) \;:=\; \int_{\underline{\omega}_1(y_1^{jk})}^{\overline{\omega}_1(y_1^{jk})} \int_{\underline{\omega}_2(y_2^{jk}|\omega_1)}^{\overline{\omega}_2(y_2^{jk}|\omega_1)} d\tilde{F}^A(\omega) \;=\; 0\,.$$

it has to be true that

$$\int_{\underline{\omega}_2(y_2^{jk},\bar{\omega}_1(y_1^{jk}))}^{\overline{\omega}_2(y_2^{jk},\bar{\omega}_1(y_1^{jk}))} r_x^0(\omega)\tilde{f}_{\omega_2}^A(\omega_2 \mid \bar{\omega}_1(y_1^{jk}))d\omega_2 \;=\; 0\,.$$

Implications of $\ell_1 \to 0$. Taking the derivative of $\Delta^{0\nu}_{x,\ell_2}(y^{jk},\ell_1,0)$ with respect to ℓ_1 and evaluating at $\ell_1 = 0$ yields

$$\Delta^{0\nu}_{x,\ell_1\ell_2}(y^{jk},0,0) \;=\; \bar{r}_x^0(y^{jk}) f_1^A(y_1^{jk})\; f_2^A(y_2^{jk} \mid \overline{\omega}_1(y_1^{jk}))\,,$$

where $f_1^A(y_1^{jk}) := \frac{d}{d\ell_1}\Big(\tilde{F}^A_{\omega_1}(\overline{\omega}_1(y_1^{jk}+\ell_1))\Big)_{|\ell_1=0}$ and $\bar{r}_x^0(y^{jk}) = r_x^0\left(\overline{\omega}_1(y_1^{jk}), \overline{\omega}_2(y_2^{jk} \mid \overline{\omega}_1(y_1^{jk}))\right)$. With

$$f^A(y^{jk}) := f_1^A(y_1^{jk})\; f_2^A(y_2^{jk} \mid \overline{\omega}_1(y_1^{jk})) \tag{94}$$

this can be written more concisely as

$$\Delta^{0\nu}_{x,\ell_1\ell_2}(y^{jk},0,0) \;=\; \bar{r}_x^0(y^{jk}) f^A(y^{jk})\,,$$

and this also proves that $\bar{\mathcal{R}}_x(y^{jk}) = \Delta^{0\nu}_{x,\ell_2\ell_1}(y^{jk},0,0)$.

H.3 Characterizing Δ^0_δ

The parameter δ is not part of the tax function T^n. Consequently, changes of δ affect tax revenue only via behavioral responses. In terms of its behavioral implications, a marginal change of δ is equivalent to a marginal change of the local intercept b^{jk} in every box B^{jk}. Therefore,

$$\begin{aligned}\Delta^0_\delta \;&=\; \textstyle\sum_{j,k} \mathbf{E}_{\Omega'}\left[\mathbf{1}(y_d^0(\omega) \in B^{jk})\left(\beta^0_{I,b}(\omega) + \beta^0_{E,b}(\omega)\right)\right] \\ &=\; \mathbf{E}_y\left[\bar{\beta}^0_{I,b}(y) + \bar{\beta}^0_{E,b}(y)\right]\,.\end{aligned}$$

H.4 Implications of $u(c) = c$

Proof that $\beta^0_{I,b}(\omega) = 0$, for all ω. We only provide a proof for a type ω so that $y^*_d(\cdot)$ satisfies first order conditions. If there is strict bunching in y_1 and y_2 there are no behavioral responses. The argument below easily generalizes to the case in which there is strict bunching in one dimension and a first order condition in the other.

We can interpret $y^*_d(\cdot)$ as a function of ω, and three parameters henceforth denoted by $\bar{b}^{jk}$, $\bar{\tau}^{jk}_1$ and $\bar{\tau}^{jk}_2$. The first order conditions are

$$u'\Big(c(y^*_d(\cdot))\Big)\Big(1 - T_1(y^*_d(\cdot)) - \bar{\tau}^{jk}_1\Big) \quad = \quad k_{1,y}(y^*_{d1}(\cdot), \omega_1) \tag{95}$$

and

$$u'\Big(c(y^*_d(\cdot))\Big)\Big(1 - T_2(y^*_d(\cdot)) - \bar{\tau}^{jk}_2\Big) \quad = \quad k_{2,y}(y^*_{d2}(\cdot), \omega_2)\ . \tag{96}$$

Let $x \in \{\bar{b}^{jk}, \bar{\tau}^{jk}_1, \bar{\tau}^{jk}_2\}$. Differentiating these two equations with respect to x and evaluating the resulting expressions at the status quo. i.e. at $\bar{b}^{jk} = \bar{\tau}^{jk}_1 = \bar{\tau}^{jk}_2 = 0$ yields the following two equations:

$$\begin{aligned}
&y^*_{d1,x}(\cdot)\left[u''(\cdot)\big(1 - T_1(\cdot)\big)^2 - u'(\cdot)\, T_{11}(\cdot) - k_{1,yy}(\cdot)\right] \\
&+ y^*_{d2,x}(\cdot)\left[u''(\cdot)\big(1 - T_1(\cdot)\big)\big(1 - T_2(\cdot)\big) - u'(\cdot)\, T_{12}(\cdot)\right] \\
&+ u''(\cdot)\, \mathbf{1}(x = \bar{b}^{jk}) + \left[u''(\cdot)\big(y^*_{d1}(\cdot) - y^{jk}_1\big) - u'(\cdot)\right]\, \mathbf{1}(x = \bar{\tau}^{jk}_1) \\
&+ u''(\cdot)\big(y^*_{d2}(\cdot) - y^{jk}_2\big)\, \mathbf{1}(x = \bar{\tau}^{jk}_2) \\
&= 0,
\end{aligned} \tag{97}$$

and

$$\begin{aligned}
&y^*_{d2,x}(\cdot)\left[u''(\cdot)\big(1 - T_2(\cdot)\big)^2 - u'(\cdot)\, T_{22}(\cdot) - k_{2,yy}(\cdot)\right] \\
&+ y^*_{d1,x}(\cdot)\left[u''(\cdot)\big(1 - T_1(\cdot)\big)\big(1 - T_2(\cdot)\big) - u'(\cdot)\, T_{21}(\cdot)\right] \\
&+ u''(\cdot)\, \mathbf{1}(x = \bar{b}^{jk}) + \left[u''(\cdot)\big(y^*_{d2}(\cdot) - y^{jk}_2\big) - u'(\cdot)\right]\, \mathbf{1}(x = \bar{\tau}^{jk}_2) \\
&+ u''(\cdot)\big(y^*_{d1}(\cdot) - y^{jk}_1\big)\, \mathbf{1}(x = \bar{\tau}^{jk}_1) \\
&= 0.
\end{aligned} \tag{98}$$

These are two equations in two unknowns and we can solve explicitly for $y^*_{d1,x}$ and $y^*_{d2,x}$. Now let $u(c) = c$, $\mathbf{1}(x = \bar{b}^{jk}) = 1$ and $\mathbf{1}(x = \bar{\tau}_1^{jk}) = \mathbf{1}(x = \bar{\tau}_2^{jk}) = 0$. Then the system becomes

$$y^*_{d1,x}(\cdot)\,[T_{11}(\cdot) + k_{1,yy}(\cdot)] + \; y^*_{d2,x}(\cdot)\,T_{12}(\cdot) \quad = \quad 0, \tag{99}$$

and

$$y^*_{d2,x}(\cdot)\,[T_{22}(\cdot) + k_{2,yy}(\cdot)] + \; y^*_{d1,x}(\cdot)\,T_{21}(\cdot) \quad = \quad 0. \tag{100}$$

Obviously, $y^*_{d1,x} = y^*_{d2,x} = 0$ is a solution and, generically, it also is the only solution to this system.

Proof that $\mathbf{E}_\omega\left[\beta^0_{E,b}(\omega)\right] = 0$. For a general reform direction h, the fraction of type ω-agents choosing earnings from both activities, only one activity, or no earnings at all can be written as functions of payoff differences. E.g, $\sigma^{12}(s,h,\omega)$ can be written as a function of $v^{12}(s,h,\omega) - v^1(s,h,\omega_1)$, $v^{12}(s,h,\omega) - v^2(s,h,\omega_2)$, and $v^{12}(s,h,\omega) - v^0(\delta)$. Likewise, $\sigma^1(s,h,\omega)$ can be written as a function of $v^1(s,h,\omega_1) - v^{12}(s,h,\omega)$, $v^1(s,h,\omega_1) - v^2(s,h,\omega_2)$, and $v^1(s,h,\omega_1) - v^0(\delta)$, etc. How are these payoff differences affected if δ changes? Consider e.g. $v^{12}(s,h,\omega) - v^1(s,h,\omega_1)$. By the envelope theorem

$$\begin{aligned}
\tfrac{\partial}{\partial\delta}\Big(v^{12}(s,h,\omega) - v^1(s,h,\omega_1)\Big) \;&=\; \tfrac{\partial}{\partial\delta}v^{12}(s,h,\omega) \;-\; \tfrac{\partial}{\partial\delta}v^1(s,h,\omega_1) \\
&=\; u'(c^0_d(\omega)) \;-\; u'(c^0_{1s}(\omega)) \\
&=\; 0\,,
\end{aligned}$$

where the last equality uses that $u'(c) = 1$, for all c. Thus, with quasi-linear in consumption preferences a change of the intercept of the consumption schedule is without effect on extensive margin choices. Since a marginal change in δ is in terms of its impact on diposable incomes equivalent to a marginal change of b^{jk} for every box B^{jk} the following statement is true when $u(c) = c$: $\mathbf{1}(y^0_d(\omega) \in B^{jk}) = 1$ for all

jk implies

$$\begin{aligned}\sum_{j,k} \mathbf{E}_{\Omega'}\left[\mathbf{1}(y_d^0(\omega) \in B^{jk})\beta_{E,b}^0(\omega)\right] &= \mathbf{E}_{\Omega'}\left[\beta_{E,b}^0(\omega)\right] \\ &= \mathbf{E}_y\left[\beta_{I,b}^0(\bar{\omega}(y)) + \beta_{E,b}^0(\bar{\omega}(y))\right] \\ &= \mathbf{E}_y\left[\bar{\beta}_{E,b}^0(y)\right] \ . \\ &= 0 \ .\end{aligned}$$

I Computations for Section 4.1

We start from Proposition 3. From our definition of the reform, using that, for every y_2, $h^j(y) = \int_{y_{1s}}^{y_1} h_1^j(s_1, y_2) ds_1$ and an integration by parts, we find that

$$\begin{aligned}(1 - \Delta_\delta^0)\, R_s(0, h^j) &= \int_{y_{2s}}^{y_{2s}+\ell_2} \int_{y_{1s}}^{\bar{y}_1} \Big(h^j(y)\, \bar{\mathcal{R}}_b(y) + h_1^j(y)\, \bar{\mathcal{R}}_1(y) \Big) dy_1\, dy_2 \\ &= \int_{y_{2s}}^{y_{2s}+\ell_2} \int_{y_{1s}}^{y_{1s}+\ell_1} h_1^j(y_1, y_2) \Big(\bar{\mathcal{R}}_1(y_1, y_2) + \textstyle\int_{y_1}^{\bar{y}_1} \bar{\mathcal{R}}_b(s_1, y_2) ds_1 \Big) dy_1\, dy_2 \\ &=: \int_{y_{2s}}^{y_{2s}+\ell_2} \int_{y_{1s}}^{y_{1s}+\ell_1} h_1^j(y_1, y_2)\, \mathbf{R}_1(y_1 \mid y_2)\, dy_1\, dy_2\end{aligned}$$

This expression can also be written as

$$(1 - \Delta_\delta^0)\, R_s(0, h^j) = \tau_1^j\ D_1^R(y_s, \ell) \tag{101}$$

where

$$D_1^R(y_s, \ell) := \int_{y_{2s}}^{y_{2s}+\ell_2} \int_{y_{1s}}^{y_{1s}+\ell_1} \mathbf{R}_1(y_1 \mid y_2)\, dy_1\, dy_2 \ .$$

J Proof of Corollary 1

J.1 Revenue

From Proposition 3,

$$(1 - \Delta_\delta^0)\, R_s(0, h) = \int_{\mathcal{Y}} \Big(h(y)\, \bar{\mathcal{R}}_b(y) + h_1(y)\, \bar{\mathcal{R}}_1(y) + h_2(y)\, \bar{\mathcal{R}}_2(y) \Big) dy \ . \tag{102}$$

Step 1. Using that $h_1(y) = h_1(y_1, 0) + \int_0^{y_2} h_{12}(y_1, s_2)ds_2$ and an integration by parts yields

$$\begin{aligned}\int_0^{\bar{y}_1}\int_0^{\bar{y}_2} h_1(y)\,\bar{\mathcal{R}}_1(y)\,dy_2\,dy_1 &= \int_0^{\bar{y}_1}\int_0^{\bar{y}_2} h_{12}(y)\int_{y_2}^{\bar{y}_2}\bar{\mathcal{R}}_1(y_1,s_2)ds_2\;\;dy_2\,dy_1 \\ &\quad + \int_0^{\bar{y}_1}\int_0^{\bar{y}_2} h_1(y_1,0)\,\bar{\mathcal{R}}_1(y_1,y_2)\,dy_2\,dy_1\,.\end{aligned} \tag{103}$$

Similarly, using that $h_2(y) = h_2(0, y_2) + \int_0^{y_1} h_{21}(s_1, y_2)ds_1$ and an integration by parts yields

$$\begin{aligned}\int_0^{\bar{y}_1}\int_0^{\bar{y}_2} h_2(y)\,\bar{\mathcal{R}}_2(y)\,dy_2\,dy_1 &= \int_0^{\bar{y}_1}\int_0^{\bar{y}_2} h_{21}(y)\int_{y_1}^{\bar{y}_1}\bar{\mathcal{R}}_2(s_1,y_2)ds_1\;\;dy_2\,dy_1 \\ &\quad + \int_0^{\bar{y}_1}\int_0^{\bar{y}_2} h_2(0,y_2)\,\bar{\mathcal{R}}_2(y_1,y_2)\,dy_1\,dy_2\,.\end{aligned} \tag{104}$$

Step 2. We rewrite $\int_{\mathcal{Y}} h(y)\,\bar{\mathcal{R}}_b(y)\,dy$ using that

$$h(y) = \int_0^{y_1} h_1(s_1,0)\,ds_1 + \int_0^{y_2} h_2(0,s_2)\,ds_2 + \int_0^{y_2}\int_0^{y_1} h_{12}(s_1,s_2)\,ds_1\,ds_2\,. \tag{105}$$

Hence,

$$\begin{aligned}\int_0^{\bar{y}_1}\int_0^{\bar{y}_2} h(y)\,\bar{\mathcal{R}}_b(y)\,dy &= \int_0^{\bar{y}_1}\int_0^{\bar{y}_2}\Big(\int_0^{y_1} h_1(s_1,0)\,ds_1\Big)\bar{\mathcal{R}}_b(y)\,dy \\ &\quad + \int_0^{\bar{y}_1}\int_0^{\bar{y}_2}\Big(\int_0^{y_2} h_2(0,s_2)\,ds_2\Big)\bar{\mathcal{R}}_b(y)\,dy \\ &\quad + \int_0^{\bar{y}_1}\int_0^{\bar{y}_2}\Big(\int_0^{y_2}\int_0^{y_1} h_{12}(s_1,s_2)\,ds_1\,ds_2\Big)\bar{\mathcal{R}}_b(y)\,dy\,.\end{aligned} \tag{106}$$

An integration by parts yields

$$\int_0^{\bar{y}_1}\int_0^{\bar{y}_2}\Big(\int_0^{y_1} h_1(s_1,0)\,ds_1\Big)\bar{\mathcal{R}}_b(y)\,dy = \int_0^{\bar{y}_1}\int_0^{\bar{y}_2} h_1(y_1,0)\Big(\int_{y_1}^{\bar{y}_1}\bar{\mathcal{R}}_b(s_1,y_2)\,ds_1\Big)dy\,, \tag{107}$$

and, similarly,

$$\int_0^{\bar{y}_1}\int_0^{\bar{y}_2}\Big(\int_0^{y_2} h_2(0,s_2)\,ds_2\Big)\bar{\mathcal{R}}_b(y)\,dy \;=\; \int_0^{\bar{y}_1}\int_0^{\bar{y}_2} h_2(0,y_2)\Big(\int_{y_2}^{\bar{y}_2}\bar{\mathcal{R}}_b(y_1,s_2)\,ds_2\Big)dy\,. \quad (108)$$

Step 3. Intermediate summary: Upon collecting terms, we find that

$$\begin{aligned}(1-\Delta_\delta^0)\,R_s(0,h) \;=\;& \int_0^{\bar{y}_1}\int_0^{\bar{y}_2} h_1(y_1,0)\Big(\bar{\mathcal{R}}_1(y_1,y_2)+\int_{y_1}^{\bar{y}_1}\bar{\mathcal{R}}_b(s_1,y_2)\,ds_1\Big)dy_2\;dy_1\\ &+\int_0^{\bar{y}_1}\int_0^{\bar{y}_2} h_2(0,y_2)\Big(\bar{\mathcal{R}}_2(y_1,y_2)+\int_{y_2}^{\bar{y}_2}\bar{\mathcal{R}}_b(y_1,s_2)\,ds_2\Big)dy_2\;dy_1\\ &+\int_0^{\bar{y}_1}\int_0^{\bar{y}_2} h_{12}(y)\int_{y_2}^{\bar{y}_2}\bar{\mathcal{R}}_1(y_1,s_2)ds_2\;\;dy_2\;dy_1\\ &+\int_0^{\bar{y}_1}\int_0^{\bar{y}_2} h_{21}(y)\int_{y_1}^{\bar{y}_1}\bar{\mathcal{R}}_2(s_1,y_2)ds_1\;\;dy_2\;dy_1\\ &+\int_0^{\bar{y}_1}\int_0^{\bar{y}_2}\Big(\int_0^{y_2}\int_0^{y_1} h_{12}(s_1,s_2)\,ds_1\,ds_2\Big)\bar{\mathcal{R}}_b(y)\;dy_2\;dy_1\,.\end{aligned} \quad (109)$$

Using the definition of the unconditional revenue functions, stated in the main text, this can equivalently be written as

$$\begin{aligned}(1-\Delta_\delta^0)\,R_s(0,h) \;=\;& \int_0^{\bar{y}_1} h_1(y_1,0)\;\bar{\mathbf{R}}_1(y_1)dy_1+\int_0^{\bar{y}_2} h_2(0,y_2)\bar{\mathbf{R}}_2(y_2)dy_2\\ &+\int_0^{\bar{y}_1}\int_0^{\bar{y}_2} h_{12}(y)\int_{y_2}^{\bar{y}_2}\bar{\mathcal{R}}_1(y_1,s_2)ds_2\;\;dy_2\;dy_1\\ &+\int_0^{\bar{y}_1}\int_0^{\bar{y}_2} h_{21}(y)\int_{y_1}^{\bar{y}_1}\bar{\mathcal{R}}_2(s_1,y_2)ds_1\;\;dy_2\;dy_1\\ &+\int_0^{\bar{y}_1}\int_0^{\bar{y}_2}\Big(\int_0^{y_2}\int_0^{y_1} h_{12}(s_1,s_2)\,ds_1\,ds_2\Big)\bar{\mathcal{R}}_b(y)\;dy_2\;dy_1\,.\end{aligned} \quad (110)$$

Step 4. We now verify that

$$\begin{aligned}&\int_0^{\bar{y}_1}\int_0^{\bar{y}_2}\Big(\int_0^{y_2}\int_0^{y_1} h_{12}(s_1,s_2)\,ds_1\,ds_2\Big)\bar{\mathcal{R}}_b(y)\;dy_2\;dy_1\\ &=\int_0^{\bar{y}_1}\int_0^{\bar{y}_2} h_{12}(y_1,y_2)\Big(\int_{y_1}^{\bar{y}_1}\int_{y_2}^{\bar{y}_2}\bar{\mathcal{R}}_b(s_1,s_2)ds_2ds_1\Big)dy_2\;dy_1.\end{aligned} \quad (111)$$

To see this, we rewrite the left-hand side of (111) as

$$\int_0^{\bar{y}_1}\int_0^{y_1}\left(\int_0^{\bar{y}_2}\int_0^{y_2}\bar{\mathcal{R}}_b(y)\; h_{12}(s_1,s_2)\; ds_2dy_2\right) ds_1dy_1.$$

The integral in brackets can be rewritten using an integration by part, so that we ultimately obtain

$$\begin{aligned}
&\int_0^{\bar{y}_1}\int_0^{\bar{y}_2}\Big(\int_0^{y_2}\int_0^{y_1} h_{12}(s_1,s_2)\, ds_1\, ds_2\Big)\bar{\mathcal{R}}_b(y)\; dy_2\; dy_1\\
&= \int_0^{\bar{y}_1}\int_0^{\bar{y}_2}\int_0^{y_1} h_{12}(s_1,y_2)\Big(\int_{y_2}^{\bar{y}_2}\bar{\mathcal{R}}_b(y_1,s_2)ds_2\Big)ds_1\; dy_2\; dy_1 \qquad (112)\\
&= \int_0^{\bar{y}_2}\left[\int_0^{\bar{y}_1}\Big(\int_{y_2}^{\bar{y}_2}\bar{\mathcal{R}}_b(y_1,s_2)ds_2\Big)\int_0^{y_1} h_{12}(s_1,y_2)ds_1\; dy_1\right]\; dy_2.
\end{aligned}$$

The term in square brackets can, after another integration by parts be written as

$$\int_0^{\bar{y}_1} h_{12}(y_1,y_2)\Big(\int_{y_1}^{\bar{y}_1}\int_{y_2}^{\bar{y}_2}\bar{\mathcal{R}}_b(y_1,s_2)ds_2ds_1\Big)dy_1$$

Reordering terms yields the expression on the right-hand side of (111).

Step 5. Using (110), (111) and that, for all interior y, $h_{12}(y) = h_{21}(y)$, allows us to rewrite $R_s(0,h)$ as

$$\begin{aligned}
(1-\Delta_\delta^0)\, R_s(0,h) \;&=\; \int_0^{\bar{y}_1} h_1(y_1,0)\;\bar{\mathbf{R}}_1(y_1)dy_1 + \int_0^{\bar{y}_2} h_2(0,y_2)\bar{\mathbf{R}}_2(y_2)dy_2\\
&\quad + \int_0^{\bar{y}_1}\int_0^{\bar{y}_2} h_{12}(y)\;\mathbf{R}^J(y)\; dy_2\; dy_1\;.
\end{aligned} \qquad (113)$$

where

$$\begin{aligned}
\mathbf{R}^J(y) \;&=\; \int_{y_2}^{\bar{y}_2}\bar{\mathcal{R}}_1(y_1,s_2)ds_2 \;+\; \int_{y_1}^{\bar{y}_1}\bar{\mathcal{R}}_2(s_1,y_2)ds_1\\
&\quad +\; \int_{y_1}^{\bar{y}_1}\int_{y_2}^{\bar{y}_2}\bar{\mathcal{R}}_b(y_1,s_2)ds_2ds_1
\end{aligned} \qquad (114)$$

J.2 Welfare

The functional derivative $W_s(0,h)$ can be computed using that

$$W_s(0,h) \;=\; R_s(0,h) - \textstyle\int_{\mathcal{Y}} h(y)\bar{g}(y)f(y)dy\ .$$

The computation of the integral on the right-hand side can proceed along the same lines as the computation of $\int_0^{\bar{y}_1}\int_0^{\bar{y}_2} h(y)\ \bar{\mathcal{R}}_b(y)\ dy_2\ dy_1$ in Section J.1, substituting $\bar{g}(y)f(y)$ for $\bar{\mathcal{R}}_b(y)$. This yields the expressions that are stated in the body of the text.

K Implications of joint taxation, piecewise linear taxes, iso-elastic effort costs and linear consumption utility

In this part of the Supplementary material we document the implications of specific functional from assumptions for our general framework. Our quantitative analysis relies on these assumptions.

Proposition 3 implies that with $u(c) = c$, we have $\bar{\beta}^0_{I,b}(y) = 0$, for all y. Therefore, by Lemma 2, we seek to characterize $\bar{\mathcal{R}}_1(y) = f^y(y)\,\bar{\beta}^0_{I,1}(y)$, $\bar{\mathcal{R}}_2(y) = f^y(y)\,\bar{\beta}^0_{I,2}(y)$ and $\bar{\mathcal{R}}_b(y) = f^y(y)\left(\bar{\beta}^0_{E,b}(y) + \bar{\mu}^0_b(y)\right)$. We recall from Appendix G.2 that

$$\begin{aligned}
\bar{\beta}^0_{I,1}(y) \;&=\; \sigma^{12,0}(\bar{\omega}(y))\left[T_1(y^0_d(\bar{\omega}(y)))\ y^0_{d1,1}(\bar{\omega}(y)) + T_2(y^0_d(\bar{\omega}(y)))y^0_{d2,1}(\bar{\omega}(y))\right]\\
\bar{\beta}^0_{I,2}(y) \;&=\; \sigma^{12,0}(\bar{\omega}(y))\left[T_1(y^0_d(\bar{\omega}(y)))\ y^0_{d1,2}(\bar{\omega}(y)) + T_2(y^0_d(\bar{\omega}(y)))y^0_{d2,2}(\bar{\omega}(y))\right]
\end{aligned}$$

$$\begin{aligned}
\bar{\beta}^0_{E,b}(y) \;=\; &\sigma_t^{12,0}(\omega)\ T(y^0_d(\bar{\omega}(y))) + \sigma_{12}^{1,0}(\bar{\omega}(y))\ T(y^0_{1s}(\bar{\omega}_1(y)),0)\\
&+\sigma_{12}^{2,0}(\bar{\omega}(y))\ T(0,y^0_{2s}(\bar{\omega}_2(y)))
\end{aligned} \tag{115}$$

where

$$\sigma_t^{12,0}(\bar{\omega}(y)) \quad = \quad \sigma_1^{12,0}(\bar{\omega}(y)) + \sigma_2^{12,0}(\bar{\omega}(y)) + \sigma_0^{12,0}(\bar{\omega}(y)) \; ,$$

and, finally,

$$\bar{\mu}_b^0(y) = \sigma^{12,0}(\bar{\omega}(y)) \; .$$

Moreover, for $u(c) = c$, differentiating Equations (95) and (96) yields

$$-y_{d1,x}^*(\cdot)\,[T_{11}(\cdot) + k_{1,yy}(\cdot)] - \; y_{d2,x}^*(\cdot)\,T_{12}(\cdot) - \; \mathbf{1}(x = \bar{\tau}_1^{jk}) \; = \; 0 \; , \tag{116}$$

and

$$-y_{d2,x}^*(\cdot)\,[\,T_{22}(\cdot) + k_{2,yy}(\cdot)] - \; y_{d1,x}^*(\cdot)\,T_{21}(\cdot) - \; \mathbf{1}(x = \bar{\tau}_2^{jk}) \; = \; 0 \; . \tag{117}$$

Imposing that $\mathbf{1}(x = \bar{\tau}_1^{jk}) = 1$ and $\mathbf{1}(x = \bar{\tau}_2^{jk}) = 0$ yields

$$y_{d1,1}^* \quad = \quad -\frac{T_{22} + k_{2,yy}}{det(H)} \; , \tag{118}$$

and

$$y_{d1,2}^* \quad = \quad \frac{T_{12}(\cdot)}{det(H)} \; , \tag{119}$$

where

$$det(H) \quad = \quad [T_{11}(\cdot) + k_{1,yy}(\cdot)]\,[\,T_{22}(\cdot) + k_{2,yy}(\cdot)] - [T_{12}]^2$$

is the determinant of the Hessian matrix that follows from the objective of the utility-maximization problem. For an interior solution that is characterized by first order conditions we need to have $det(H) > 0$. The following Lemma states an immediate implication of these conditions: With positive jointness, primary and secondary earnings move in opposite directions when the marginal tax on primary earnings is changed.

Lemma 3 *If y_d^* is characterized by first order conditions, then $y_{d1,1}^*(\cdot) < 0$. With $T_{12}(\cdot) > 0$ (positive jointness), we have $y_{d2,1}^*(\cdot) > 0$.*

Plugging the expressions in (116) and (117) into the formula for $\bar{\beta}_{I,1}^0(y)$ yields

$$\bar{\beta}_{I,1}^0(y) = -\frac{\sigma^{12,0}(\cdot)}{det(H)} \left[T_1(\cdot)(T_{22}(\cdot) + k_{2,yy}(\cdot)) - T_2(\cdot)\, T_{12}(\cdot) \right] \tag{120}$$

Joint taxation implies that $T_1(\cdot) = T_2(\cdot)$ and $T_{12}(\cdot) = T_{22}(\cdot)$. Under these premises (120) can be rewritten as

$$\bar{\beta}_{I,1}^0(y) = -\frac{\sigma^{12,0}(\cdot)}{det(H)}\, T_1(\cdot)\, k_{2,yy}(\cdot) \tag{121}$$

implying that local increases of the marginal tax rate on the earnings of spouse 1 come with a substitution effect that is costly in terms of tax revenue.

The analysis of the case $\mathbf{1}(x = \bar{\tau}_1^{jk}) = 0$ and $\mathbf{1}(x = \bar{\tau}_2^{jk}) = 1$ proceeds along the same lines and yields

$$y_{d1,1}^* = -\frac{T_{12}(\cdot)}{det(H)}\,, \tag{122}$$

and

$$y_{d1,2}^* = -\frac{T_{11}(\cdot) + k_{1,yy}(\cdot)}{det(H)}\,, \tag{123}$$

and

$$\bar{\beta}_{I,2}^0(y) = -\frac{\sigma^{12,0}(\cdot)}{det(H)} \left[T_2(\cdot)(T_{11}(\cdot) + k_{1,yy}(\cdot)) - T_1(\cdot)\, T_{12}(\cdot) \right]. \tag{124}$$

If the status quo has joint taxation then this can be written as

$$\bar{\beta}_{I,2}^0(y) = -\frac{\sigma^{12,0}(\cdot)}{det(H)} T_2(\cdot)\, k_{1,yy}(\cdot)\,. \tag{125}$$

Iso-elastic effort cost functions. Suppose that the variable effort cost functions k_1 and k_2 are iso-elastic; i.e.

$$k_1(y_1, \omega_1) \;=\; \frac{1}{1+\frac{1}{\varepsilon_1}} \left(\frac{y_1}{\omega_1}\right)^{1+\frac{1}{\varepsilon_1}} \quad \text{and} \quad k_2(y_2, \omega_2) \;=\; \frac{1}{1+\frac{1}{\varepsilon_2}} \left(\frac{y_2}{\omega_2}\right)^{1+\frac{1}{\varepsilon_2}} .$$

Then, we have that, for any $i = 1, 2$,

$$k_{i,yy}(\cdot) \;=\; \frac{1}{\varepsilon_i}\,\frac{1}{y_i}\,k_{i,y}(\cdot)$$

When first order conditions are satisfied, we have, moreover,

$$k_{i,y}(\cdot) \;=\; 1 - T_i(\cdot)$$

and, therefore,

$$k_{i,yy}(\cdot) \;=\; \frac{1}{\varepsilon_i}\,\frac{1}{y_i}\,\Big(1 - T_i(\cdot)\Big) \tag{126}$$

Define the coefficients of progressivity

$$\rho^1(y) \;=\; -\frac{T_{11}(\cdot)}{1 - T_1(\cdot)}\, y_1 \;,$$

$$\rho^2(y) \;=\; -\frac{T_{22}(\cdot)}{1 - T_2(\cdot)}\, y_2 \;,$$

$$\kappa^1(y) \;=\; -\frac{k_{1,yy}(\cdot)}{k_{1,y}}\, y_1 \;,$$

and

$$\kappa^2(y) \;=\; -\frac{k_{2,yy}(\cdot)}{k_{2,y}}\, y_2 \;.$$

Also note that, with iso-elastic effort costs,

$$\kappa^1(y) \;=\; \frac{1}{\varepsilon_1} \;,$$

and

$$\kappa^2(y) \quad = \quad \frac{1}{\varepsilon_2} \, .$$

Note that, using this notation, $det(H)$, can be written as

$$det(H) \quad = \quad \frac{1-T_1(\cdot)}{y_1}\frac{1-T_2(\cdot)}{y_2}\Big(\rho^1(y)-\kappa^1(y)\Big)\Big(\rho^2(y)-\kappa^2(y)\Big) - T_{12}(\cdot)^2 \, .$$

Joint taxation implies that $T_1 = T_2$ and also $T_{11} = T_{22} = T_{12}$ so that this may also be written as

$$det(H) \quad = \quad \frac{\big(1-T_1(\cdot)\big)^2}{y_1 \; y_2}\left\{\left(\rho^1(y)-\frac{1}{\varepsilon_1}\right)\left(\rho^2(y)-\frac{1}{\varepsilon_2}\right) - \rho^1(y)\rho^2(y)\right\} \, . \quad (127)$$

The following Lemma states implications of equations (121), (125), (126) and (127).

Lemma 4 *If the status quo has joint taxation and the effort cost functions are iso-elastic, then*

$$\bar{\beta}^0_{I,1}(y) = -\frac{\sigma^{12,0}(\cdot)}{\bar{H}(y)} \; \frac{T_1(\cdot)}{1-T_1(\cdot)} \; \frac{y_1}{\varepsilon_2} \quad (128)$$

and

$$\bar{\beta}^0_{I,2}(y) = -\frac{\sigma^{12,0}(\cdot)}{\bar{H}(y)} \; \frac{T_2(\cdot)}{1-T_2(\cdot)} \; \frac{y_2}{\varepsilon_1} \quad (129)$$

where

$$\bar{H}(y) := \left(\rho^1(y)-\frac{1}{\varepsilon_1}\right)\left(\rho^2(y)-\frac{1}{\varepsilon_2}\right) - \rho^1(y)\rho^2(y) \, .$$

If the tax system is, moreover, (locally) linear then $\rho^1(y) = \rho^2(y) = 0$ and these terms simplify:

$$\bar{\beta}^0_{I,1}(y) = -\sigma^{12,0}(\cdot) \; \frac{T_1(\cdot)}{1-T_1(\cdot)} \; \frac{y_1}{\varepsilon_1} \quad (130)$$

and

$$\bar{\beta}^0_{I,2}(y) = -\sigma^{12,0}(\cdot)\,\frac{T_2(\cdot)}{1-T_2(\cdot)}\,\frac{y_2}{\varepsilon_2}\,. \tag{131}$$

K.1 Cutoff-types in Φ for given ω

A couple will choose to be a dual earner couple when the resulting payoff of $v^{12}(\omega,s,h)-\varphi_1-\varphi_2$. exceeds the payoff from its alternative options. Alternatives are that only spouse 1 or only spouse 2 exerts productive effort with payoffs that are, respectively, given by $v_1(\omega_1,s,h)-\varphi_1$ and $v_2(\omega_2,s,h)-\varphi_2$. Thus, for dual earner couples,

$$v^{12}(\omega,s,h)-v^1(\omega_1,s,h) \;:=\; \hat{\varphi}_2(\omega,s,h) \quad\geq\quad \varphi_2\,, \quad \text{and}$$

$$v^{12}(\omega,s,h)-v^2(\omega_2,s,h) \;:=\; \hat{\varphi}_1(\omega,s,h) \quad\geq\quad \varphi_1\,.$$

It also has to be true that $v^{12}(\omega,s,h)\geq\varphi_1+\varphi_2$. Otherwise, the couple would be better off with no earnings at all.

A couple will choose spouse i as its single-earner when this yields a payoff that dominates what is achievable as a dual earner couple or with spouse $-i$ as the single earner. The latter implies

$$v^i(\omega_i,s,h)-\varphi_i \quad\geq\quad v^{-i}(\omega_{-i},s,h)-\varphi_{-i}\,.$$

The payoff with i as the single earner must also dominate having no earnings at all,

$$v^i(\omega_i,s,h) \;:=\; \hat{\hat{\varphi}}_i(\omega_i,s,h) \quad\geq\quad \varphi_i\,.$$

Finally, for couples with no earnings, it must be true that, for both i, $\varphi_i\geq\hat{\hat{\varphi}}(\omega_i,s,h)$ and $v^{12}(\omega,s,h)\;\leq\;\varphi_1+\varphi_2$. The following Lemma states implications of these conditions under a set of additional assumptions, see Figure K2 for an illustration.

Lemma 5 *Suppose that the following premises are satisfied:*

i) For all i, $y^*_{si}(\omega_i,s,h)>y^*_{di}(\omega,s,h)$.

ii) For all i, $\lim_{\omega_i\to 0} v^i(\omega_i,s,h)=0$ and $\lim_{\omega_{-i}\to 0} v^{12}(\omega,s,h)=v^i(\omega_i,s,h)$.

Then, for any $\omega >> 0$*, and for any* $i = 1, 2$*,*

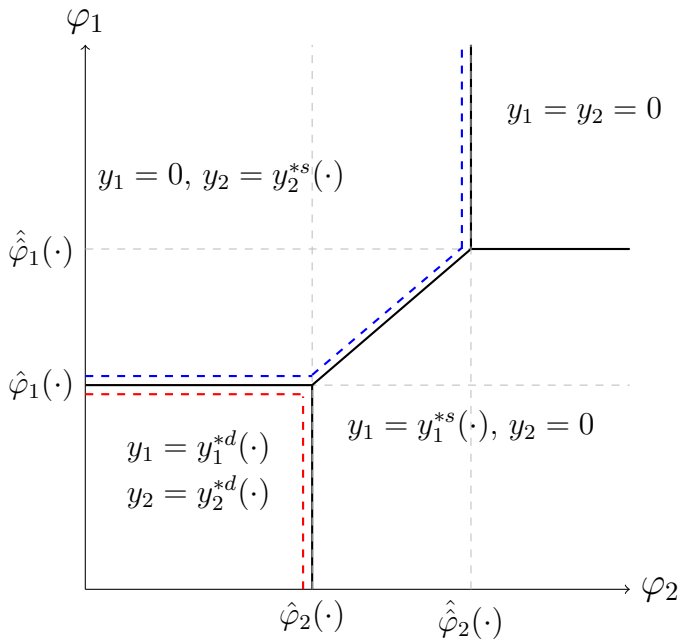
$$\hat{\hat{\varphi}}(\omega_i, \tau, h) \quad > \quad \hat{\varphi}_i(\omega, \tau, h) \; . \tag{132}$$

Hence, $y^*(\omega, \varphi, s, h) =$

$$\begin{cases} y_d^*(s,h,\omega), & \text{if} \;\; \varphi_1 < \hat{\varphi}_1(\cdot) \;\; \text{and} \;\; \varphi_2 < \hat{\varphi}_2(\cdot) \; , \\ (y_{s1}^*(s,h,\omega_1), 0), & \text{if} \;\; \varphi_1 < \min\{\Delta v^s(\cdot) + \varphi_2, \hat{\hat{\varphi}}_1(\cdot)\} \;\; \text{and} \;\; \varphi_2 > \hat{\varphi}_2(\cdot) \; , \\ (0, y_{s2}^*(s,h,\omega_2)), & \text{if} \;\; \varphi_1 > \hat{\varphi}_1(\cdot) \;\; \text{and} \;\; \varphi_2 < \min\{-\Delta v^s(\cdot) + \varphi_1, \hat{\hat{\varphi}}_2(\cdot)\} \; , \\ (0,0), & \text{if} \;\; \varphi_1 > \hat{\hat{\varphi}}_1(\cdot) \;\; \text{and} \;\; \varphi_2 > \hat{\hat{\varphi}}_2(\cdot) \; . \end{cases} \tag{133}$$

where $\Delta v^s(\omega, s, h) := v^1(\omega_1, \tau, h) - v^2(\omega_2, s, h)$.

Figure K2: Cutoff types in Φ



Note: The figure shows how the earnings choices of the spouses in a couple with productivities $\omega = (\omega_1, \omega_2)$ depend on their fixed costs $\varphi = (\varphi_1, \varphi_2)$. For fixed costs in the lower left rectangle both spouses generate income. For fixed costs in the upper right area, the couple has no earnings at all. For fixed costs in the upper left area, spouse 1 is the single earner. For fixed costs in the lower right area, spouse 2 is the single earner.

The first premise of Lemma 5 is fulfilled under a progressive tax system that gives rise to positive jointness. Under such a system, the marginal tax rate on the earnings of one spouse is increasing in the earnings of the other spouse. Thus, when spouse i is a single earner her marginal tax rate is lower as compared to any other constellation in which she is part of a dual earner couple. Therefore, y_i^* is decreasing

in y_{-i}. The second premise holds with an iso-elastic cost function.[60]

Proof of Lemma 5. We show that under the premises of the Lemma, we have, for all $\omega >> 0$.

$$\hat{\varphi}_1(\omega, s, h) \;<\; \hat{\hat{\varphi}}_1(\omega_1, s, h) \quad \text{and} \quad \hat{\varphi}_2(\omega, s, h) \;<\; \hat{\hat{\varphi}}_2(\omega_2, s, h) \; .$$

We only show that $\hat{\varphi}_1(\cdot) \;<\; \hat{\hat{\varphi}}_1$, or, equivalently, that

$$v^{12}(\omega, s, h) - v^2(\omega_2, s, h) \quad < \quad v^1(\omega_1, s, h) \; .$$

By assumption,

$$\lim_{\omega_2 \to 0} v^{12}(\omega, s, h) - v^2(\omega_2, s, h) \quad = \quad v^1(\omega_1, s, h) \; .$$

It therefore suffices to show that $v^{12}(\omega, s, h) - v^2(\omega_2, s, h)$ is a decreasing function of ω_2. An application of the Envelope Theorem reveals that

$$\frac{\partial}{\partial \omega_2} v^{12}(\omega, s, h) \;=\; -k_{2,\omega}(y^*_{d2}(\cdot), \omega_2)$$

and

$$\frac{\partial}{\partial \omega_2} v^2(\omega_2, s, h) \;=\; -k_{2,\omega}(y^*_{s2}(\cdot), \omega_2)$$

where, by assumption, $y^*_{s2}(\cdot) > y^*_{d2}(\cdot)$. Hence,

$$\begin{aligned} \tfrac{\partial}{\partial \omega_2}\Big(v^{12}(\omega, s, h) - v^2(\omega_2, s, h)\Big) \;&=\; k_{2,\omega}(y^*_{s2}(\cdot), \omega_2) - k_{2,\omega}(y^{*d}_2(\cdot), \omega_2) \\ &=\; \textstyle\int_{y^*_{d2}(\cdot)}^{y^*_{s2}(\cdot)} k_{2,y\omega}(y, \omega_2)\, dy \; , \end{aligned}$$

where $k_{2,y\omega}(y, \omega_2) < 0$ by the Spence-Mirrlees single crossing property.

60. More generally, it holds for any cost function k so that y^*_i goes to zero as ω_i goes to zero with the implication that also i's contribution to the couple's payoff goes to zero.

K.2 A characterization of $L_E(y_1, y_2)$ in terms of extensive margin elasticities.

We show that the revenue losses due to extensive margin responses can be written as a sum of extensive margin elasticities. For brevity, we focus on the revenue losses that arise when the tax burden increases in the interior of the income space so that the extensive margin responses are as indicated by the red dashed line in Figure K2. The analysis of a change of the tax burden on the boundary of the income space would follow the same route, except that the extensive margin responses are then as indicated by the blue dashed line in Figure K2.

Corollary 2 *Under the premises of Lemma 5,*

$$\begin{aligned}\sigma_t^{12,0}(\omega) &= \sigma_1^{12,0}(\bar{\omega}(y)) + \sigma_2^{12,0}(\bar{\omega}(y)) \\ &= -\sigma_{12}^{1,0}(\bar{\omega}(y)) - \sigma_{12}^{2,0}(\bar{\omega}(y)) \,.\end{aligned}$$

and therefore

$$\begin{aligned}\bar{\beta}_{E,b}^0(y) = \; & -\sigma_{12}^{1,0}(\bar{\omega}(y))\Big(T(y_d^0(\bar{\omega}(y))) - T(y_{s1}^0(\bar{\omega}_1(y)), 0)\Big) \\ & -\sigma_{12}^{2,0}(\bar{\omega}(y))\Big(T(y_d^0(\bar{\omega}(y))) - T(y_{s2}^0(0, \bar{\omega}_2(y)))\Big) \,.\end{aligned} \tag{134}$$

Hence, we can write

$$\bar{\beta}_{E,b}^0(y) + \bar{\mu}_b^0(y) = \sigma^{12,0}\Big(1 - L_E^{12,0}(y)\Big) \,, \tag{135}$$

where $L_E^{12,0}(y)$ is a loss of revenue due to extensive margin responses. This loss can be written as

$$\begin{aligned}L_E^{12,0}(y) = \; & \chi^{12,1}(y)\Big(T(y_d^0(\bar{\omega}(y))) - T(y_{s1}^0(\bar{\omega}_1(y)), 0)\Big) \\ & + \; \chi^{12,2}(y)\Big(T(y_d^0(\bar{\omega}(y))) - T(y_{s2}^0(0, \bar{\omega}_2(y)))\Big) \,.\end{aligned}$$

where $\chi^{12,1}(y) := \frac{\sigma_{12}^{1,0}}{\sigma^{12,0}}$ and $\chi^{12,2}(y) := \frac{\sigma_{12}^{2,0}}{\sigma^{12,0}}$ are semi-elasticities that relate a percentage change in dual earner couples to a marginal change in the payoff differences $v^1(s, h, \omega_1) - v^{12}(s, h, \omega)$ and $v^1(s, h, \omega_1) - v^{12}(s, h, \omega)$.

Denote the corresponding extensive margin elasticities by $\mathcal{E}_x^{12,1}(y) := \chi^{12,1}(y)(v^{12}(\cdot) - v^1(\cdot))$ and $\mathcal{E}_x^{12,2}(y) := \chi^{12,2}(y)(v^{12}(\cdot) - v^2(\cdot))$. Then, we can write

$$\begin{aligned} L_E^{12,0}(y) &= \mathcal{E}_x^{12,1}(y) \, \tfrac{T(y_d^0(\bar{\omega}(y))) - T(y_{s1}^0(\bar{\omega}_1(y)),0)}{v^{12}(\cdot) - v^1(\cdot)} \\ &+ \mathcal{E}_x^{12,2}(y) \tfrac{T(y_d^0(\bar{\omega}(y))) - T(y_{s2}^0(0,\bar{\omega}_2(y)))}{v^{12}(\cdot) - v^2(\cdot)} \, . \end{aligned}$$

For s close to 0 and a general tax perturbation h, the envelope theorem and a first order Taylor approximation imply that

$$v^{12}(\cdot) - v^2(\cdot) \quad \simeq \quad - s\Big(h(y_d^0(\cdot)) - h(0, y_{s2}^0(\cdot))\Big) \, .$$

Hence, $v^{12}(\cdot) - v^2(\cdot)$ is proportional to the absolute change in the "participation tax" $T(y_d^0(\bar{\omega}(y))) - T(y_{s2}^0(0, \bar{\omega}_2(y)))$ and

$$\frac{T(y_d^0(\bar{\omega}(y))) - T(y_{s2}^0(0, \bar{\omega}_2(y)))}{v^2(\cdot) - v^{12}(\cdot)} \quad := \quad \frac{1}{PPT^{12,1}}$$

is the inverse of a percentage change of the participation tax. We can now write

$$L_E^{12,0}(y) = \tilde{\mathcal{E}}_x^{12,1}(y) + \tilde{\mathcal{E}}_x^{12,2}(y) \, ,$$

where

$$\tilde{\mathcal{E}}_x^{12,1}(y) := \frac{\mathcal{E}_x^{12,1}(y)}{PPT^{12,1}} \quad \text{and} \quad \tilde{\mathcal{E}}_x^{12,2}(y) := \frac{\mathcal{E}_x^{12,2}(y)}{PPT^{12,2}}$$

When bringing the theory to data, we rely on estimates of extensive margin elasticities. More specifically, we interpret this literature as providing estimates of $\tilde{\mathcal{E}}_x^{12,1}(y)$, $\tilde{\mathcal{E}}_x^{12,2}(y)$, or, directly of $L_E^{12,0}(y)$.

L Appendix for the Application

This section of the appendix contains additional explanations, supplementary graphs and tables on the empirical part of the paper in Section 6. We first present details on the data preparation (Section L.1). Subsequently, supplementary insights are presented separately for each of the respective subset of results, i.e., for results on Pareto efficiency (Section L.2), tax reforms at the bottom of the distribution (Section L.3.5), welfare effects of jointness-reducing reforms (Section L.4), and optimal top tax rates (Section L.5).

L.1 Data

In Section 6 of the main text, we use household micro data from the Current Population Survey (CPS) and tabulated income tax return data from the Statistics of Income (SOI) program. CPS data is used for the estimation of conditional revenue functions (Section 6.1), tax reforms at the bottom of the income distribution (Section 6.1), as well as for analyzing the welfare effects of reducing jointness (Section 6.2). For the calibration of optimal top tax rates in Section 6.3, we use tabulated income tax return data. In the following, we describe the details of the two data sources and respective data preparation.

L.1.1 Current Population Survey (CPS)

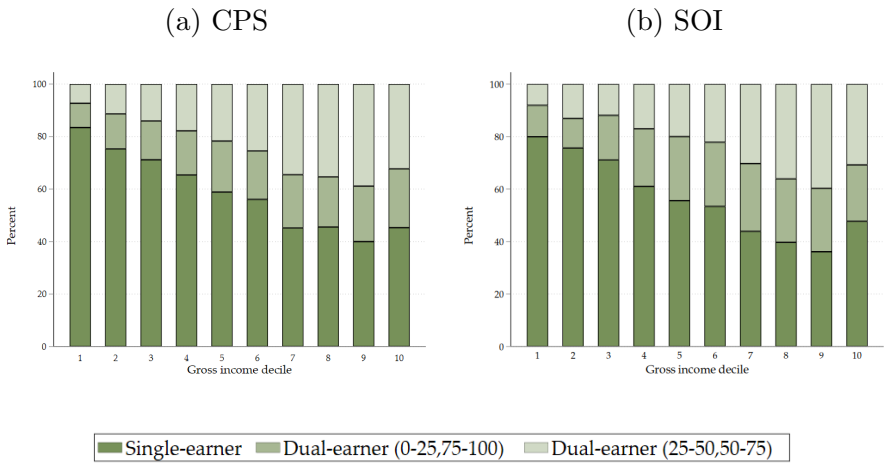
The Current Population Survey (CPS) is conducted by the US Census Bureau and the Bureau of Labor Statistics and contains nationally representative cross-sectional survey data from 1962 onward. We use data from the Annual Social and Economic Supplement of the Current Population Survey (CPS-ASEC).[61] The sample size of CPS-ASEC increased from around 30,000 households in 1962 to more than 90,000 in the most recent wave. In contrast to tax return micro data such as the public use files (IRS-SOI PUF) from the Statistics of Income (SOI) division of the Internal Revenue Service (IRS), as, e.g., used by Bargain et al. (2015) or Bierbrauer, Boyer, and Peichl (2021), the CPS data contain exact information about the incomes of primary and secondary earners of the tax unit.[62]

61. See Flood et al. (2021) and https://cps.ipums.org for a detailed description of CPS data.

62. In the IRS-SOI PUF, the relevant information on salaries and wages from the W2-form of the primary and secondary earner is only available for the year 1974 and imputed for all other

To adapt the CPS to the input requirements of the microsimulation model, we transform the CPS from a household-level data set to a tax unit level data set. For this purpose, we form tax units by joining all married spouses with their dependent children. Single individuals and unmarried spouses form separate tax units. Children of single individuals are in most cases allocated to the household head. Adult individuals with a total income below the year-specific personal exemption threshold are assumed to reflect dependents of the household head. Table L1 illustrates in detail the correspondence between variables utilized in NBER TAXSIM and variables in the CPS data.

Figure L3: Comparison of CPS and SOI data (1974), couple types



Note: This figure displays for the tax year of 1974 the distribution of married couple types across deciles of the per-capita income distribution. The figure compares the distribution based on the CPS data (Figure L3a) to the distribution based on the IRS-SOI PUF tax return micro data (Figure L3b). All estimates are based on tax units with strictly positive gross income in which both spouses are between 25 and 55 years old.
Source: Our calculations based on CPS-ASEC and SOI PUF.

Treatment of top incomes. In the CPS data, information on top incomes is limited by (i) public top coding, and (ii) internal censoring. We address both limitations by harmonizing the treatment of top incomes across the different survey years and by following Piketty and Saez (2003) and Piketty (2003) in assuming that top incomes are well represented by a Pareto distribution.

years using an undocumented procedure. For 1974, in which reliable information is available, the distribution of different couple types across per-capita income distribution is very similar to the CPS data (see Figure L3). Moreover, Bargain et al. (2015) compare inequality measures as well as the direct effects of tax policies based on CPS and SOI-PUF data and show that results are very similar (except for the very top of the distribution).

In a first step, we address the challenge that public top coding methods vary over time. In most recent years (since 2011), the Census Bureau uses a rank proximity swapping procedure to preserve the privacy for top income earners while maintaining the internal distribution of top incomes. In this procedure, values at or above a specific swap threshold are switched against other top income values within a bounded interval. For previous years, however, the CPS data originally contains top income values that are based on different procedures, in particular traditional top coding (1962-1995), and a replacement value system procedure (1996-2010). To be able to consistently analyze the data, we apply the most recent method of rank proximity swapping also to previous years using supplementary files provided by IPUMS.[63] Thereby, we preserve the internally used distribution of top incomes whenever possible.

In a second step, we address the challenge that top incomes are also internally censored based on the value range limits of the income variables. As shown by Larrimore et al. (2008), since these censoring thresholds have changed discretely at specific points in time, the share of individuals affected by censoring varies and can reach up to one percent in specific years. To address the unequal representation of censored incomes, we replace censored incomes by random draws from a Pareto distribution. In particular, we first identify for every year and every income type the highest possible income T assigned in a given year. Based on this censoring threshold, we generate for every year and every income type the parameter α of a Pareto distribution with density $f(Y) = \alpha * T^{\alpha} * Y^{-\alpha-1}$. We thereby assume that incomes above the 99th percentile follow a Pareto distribution and thus estimate the shape parameter α as

$$\alpha = \frac{ln\left(\frac{N_{Y \geq p99}}{N_{Y=T}}\right)}{ln\left(\frac{Y_T}{Y_{p99}}\right)}$$

where $N_{Y \geq p99}$ is the number of individuals with an income above the 99th percentile of the income distribution, $N_{Y=T}$ is the number of individuals at the highest income, and Y_T and Y_{p99} are the top income and the income at the 99th percentile respec-

63. For details on the treatment of top incomes in general and the data used for rank proximity swapping, see https://cps.ipums.org/cps/topcodes_tables.shtml and https://cps.ipums.org/cps/income_cell_means.shtml.

tively.[64] Finally, we use the distribution to replace the top incomes T by random draws from this calibrated distribution.[65]

Sample restrictions. In this paper, we focus on married couples.[66] We assume that married couples always file jointly. While married couples can also file separately, this filing status is usually not beneficial (see Figure L5) and is chosen by less than two percent of all tax units (see Figure L4).[67] Similarly, we abstract from the qualifying widow(er) filing status that gives widowed individuals a preferential tax treatment in the two years following the spouses' death. Given our sample restriction, the occurrence of widow(er)s is negligible (see also Figure L4). If not indicated otherwise, we restrict the sample to tax units in which primary and secondary taxpayer are between 25 and 55 years old and have non-negative gross income. This sample restriction is guided by (i) our model that considers neither education nor retirement decisions, and (ii) the assumptions on labor supply responses to taxation that are not valid for young and old people with weak labor force attachment.

Throughout the analysis, we calculate tax payments as well as average and marginal tax rates based on the federal income tax and abstract from state income tax and social security payroll taxes. Our pre-tax gross income variable of interest contains wage income, farm income, business income, income from dividends, income from interest, income from rent, and retirement income.

64. Discussions of different estimation methods for the shape parameter of the Pareto distribution can be found in Armour, Burkhauser, and Larrimore (2016) and Blanchet, Fournier, and Piketty (2022).

65. To reduce the impact of random sampling on our results, we use quantiles of the distribution. The number of quantiles utilized depends on the number of individuals at the top income. For instance, if we observe 25 individuals at the top income, we assign these individuals income levels that correspond to the 25 quantiles of the randomly drawn values from the calibrated Pareto distribution. Thereby, we preserve the properties of the distribution while limiting the influence of random draws.

66. Our companion paper Bierbrauer et al. (2024) discusses the relationship between singles and married couples explicitly.

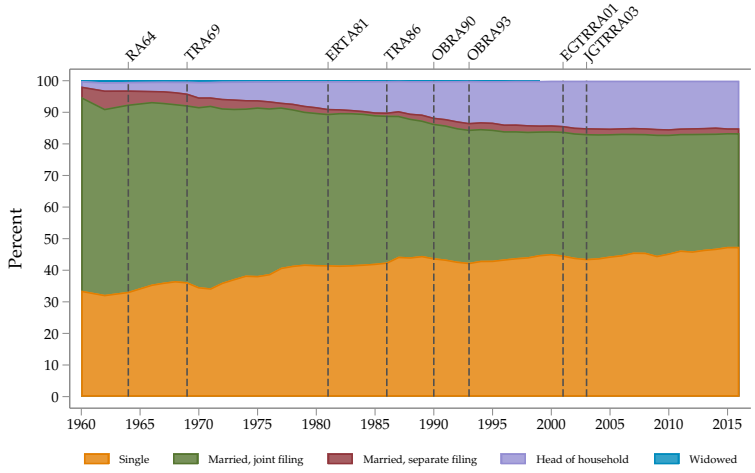
67. Filing separately can be beneficial in very particular circumstances that we do not observe, i.e., in the case of substantial itemized deductions (e.g. high medical expenses or student loan repayments).

Table L1: TAXSIM variables and CPS application

TAXSIM Variable	Explanation	CPS Application
taxsimid	Case ID	N/A
year	Tax year	ASEC income reference year
state	State	State of residence
mstat	Marital Status	Marital status (married vs. unmarried)
page	Age of primary taxpayer	Age of husband
sage	Age of spouse	Age of spouse
depx	Number of dependents	Number of children below and of age 18 + additional dependents
dep13	Number of children under 13	Number of children under 13
dep17	Number of children under 17	Number of children under 17
dep18	Number of qualifying children for EITC.	Number of children below and of age 18
pwages	Wage and salary income of Primary Taxpayer	Wage income + business income + farm income of husband
swages	Wage and salary income of Spouse	Wage income + business income + farm income of spouse
dividends	Dividend income	Income from dividends
intrec	Interest Received	Income from interest
stcg	Short Term Capital Gains or losses	N/A
ltcg	Long Term Capital Gains or losses.	Capital gains - capital losses
otherprop	Other property income	Income from rent
nonprop	Other non-property income	Income from other Source not specified + income from alimony
pensions	Taxable Pensions and IRA distributions	Retirement income
gssi	Gross Social Security Benefits	Social Security income
ui	Unemployment compensation received	Income from unemployment benefits
transfers	Other non-taxable transfer Income	Welfare (public assistance) income + income from worker's compensation + income from veteran benefits + income from survivor benefits + income from disability benefits + income from child support + income from educational assistance + income from SSI + income from assistance
rentpaid	Rent Paid	N/A
proptax	Real Estate taxes paid	Annual property taxes
otheritem	Other Itemized deductions	Indirect calculation via difference between adjusted gross income and taxable income calculated by the Census Bureau's tax model.
childcare	Child care expenses	N/A
mortgage	Deductions not included in otheritem	N/A

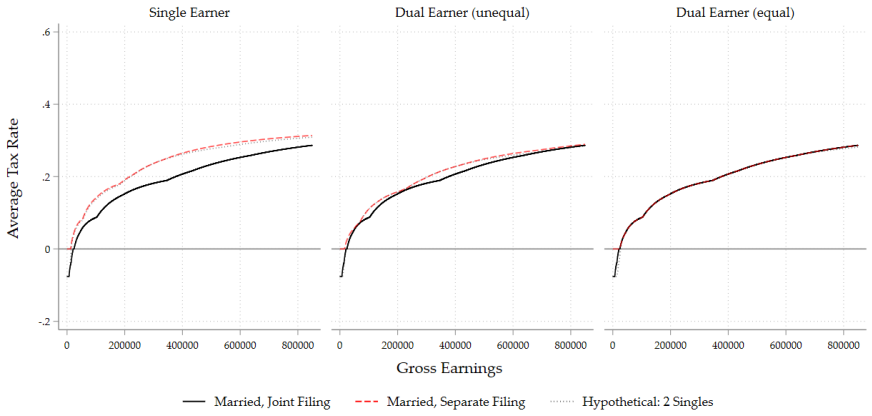
Note: This table displays the variables utilized as part of the tax calculation via the NBER TAXSIM (v32) microsimulation model and the corresponding information from the CPS used for the respective variables. For details on TAXSIM (v32) see Feenberg and Coutts (1993) and https://users.nber.org/~taxsim/.
Source: NBER TAXSIM and CPS-ASEC

Figure L4: Filing status according to SOI data



Note: This figure shows the distribution of filing status from 1960 to 2016. Filing statuses are based on the IRS-SOI PUF administrative tax return micro data.
Source: Bierbrauer et al. (2024), calculations based on SOI PUF

Figure L5: Married couples filing jointly and separately (2019)



Note: This figure shows how the average tax rate of a couple with specific gross earnings differs between whether this couple files separately or jointly. In addition, the figure also shows the average tax rate of two singles with the same joint income. The figure differentiates further by the type of couple: single earner couples (95% / 5%), unequal dual earner couples (75% / 25%) and dual earner couples with equal incomes (50% / 50%).
Source: Bierbrauer et al. (2024), calculations based on NBER TAXSIM and CPS-ASEC

L.1.2 IRS income tax tabulations

For the estimation of Pareto coefficients, we use tabular data from the Statistics of Income (SOI) program.[68] The data contains detailed tabular information on joint return taxpayers with wage income, separated by the size of wage income, gender, and the share of total wage income. It includes all primary and secondary taxpayers with wages, tips and other compensation as reported in the wage and tax statement of income tax return form W-2.

Data preparation. We prepare the tables for 2008-2018 in two steps. First, since some cells have been merged by the IRS to avoid information disclosure for specific taxpayers, we need to impute some information. Such an imputation is not necessary to estimate Pareto coefficients for a particular group. However, to aggregate information across groups, e.g., to obtain data on all secondary earners within a specific income range, the income ranges need to be harmonized. To fill in missing values for unmerged cells, we distribute the merged number of taxpayers and wages according to the distribution of all taxpayers of a specific gender. For instance, for the tax year 2018, among women with a wage share between 25 and 50 percent, there are 127 taxpayers who earn between 1.5 million and 10 million (merged cell). To allocate these taxpayers to the more detailed income brackets of 1.5-2 / 2-5 / 5-10 million (unmerged cells), we use information on how taxpayers are distributed within the merged income cell among all women. This allocation procedure is applied for both the number of taxpayers and the sum of wages, separately for men and women.

Second, we combine the cell information for men and women with different wage shares to data on primary and secondary earners, i.e. we combine the number of taxpayers and the sum of wages for men and women with wage shares between 0 and 50 percent (secondary earners) and for men and women with wage shares between 50 and 100 percent (primary earners). The final table contains separate primary and secondary earner data on the number of taxpayers and the sum of wages for different income ranges. We further produce tables for single earner couples and equal earning dual earner couples, for which the primary (secondary) earner's income share lies between 50 (25) and 75 (50) percent. In case of inconsistencies,

68. See IRS, Statistics of Income Division, Form W-2 study, June 2019.

for instance when the imputation yields an average income that is lower than the bracket threshold, we abstract from this data point for the Pareto interpolation.

Pareto interpolation. We use the tabular information on the number of taxpayers and the sum of wage income in a particular wage income bracket to generate empirical quantiles associated with the lower threshold and average wage income of a specific wage income bracket. We then follow the generalized Pareto interpolation technique by Blanchet, Fournier, and Piketty (2022) to obtain information on specific percentiles of the distribution and the associated Pareto coefficient. In contrast to standard Pareto interpolation, generalized Pareto interpolation in principle allows the Pareto coefficient to vary across the distribution and thereby provide a more flexible way to describe distributions based on tabular data.

L.2 Estimating the Effect of Tax Reforms

In Section 6 of the main text, we evaluate the revenue and welfare consequences of different types of tax reforms moving towards individual taxation. Here, we first explain, how the ingredients of revenue functions are constructed. In the subsequent sections of this Appendix, we then provide details and robustness analyses for the reform types covered in the analysis, in particular for reductions in secondary earner marginal tax rates (Section L.3), changing the jointness of the tax system (Section L.4), and for optimal top tax rates (Section L.5).

From theory to empirics Under the assumptions discussed at the beginning of Section 6, the following Lemma describes what the functions $y \mapsto \bar{\mathcal{R}}_1(y)$, $y \mapsto \bar{\mathcal{R}}_2(y)$, and $y \mapsto \bar{\mathcal{R}}_b(y)$ look like.[69] The statement of the Lemma invokes some notation. It uses

$$\rho^1(y) \quad = \quad -\frac{T_{11}(\cdot)}{1-T_1(\cdot)}\, y_1 \quad \text{and} \quad \rho^2(y) \quad = \quad -\frac{T_{22}(\cdot)}{1-T_2(\cdot)}\, y_2 \ ,$$

as local measures of the tax system's progressivity. It also invokes

$$\bar{H}(y) := \left(\rho^1(y) - \frac{1}{\varepsilon_1}\right)\left(\rho^2(y) - \frac{1}{\varepsilon_2}\right) - \rho^1(y)\rho^2(y) \ ,$$

which is "essentially" the determinant of a Hessian matrix.[70]

Lemma 6 *Suppose that $u(c) = c$ and that effort costs are iso-elastic. Suppose that the status quo has joint taxation. Then $\bar{\mathcal{R}}_1(y) = \bar{\beta}^0_{I,1}(y)f(y)$ and $\bar{\mathcal{R}}_2(y) = \bar{\beta}^0_{I,2}(y)f(y)$ where*

$$\bar{\beta}^0_{I,1}(y) = -\frac{\sigma^{12,0}(\cdot)}{\bar{H}(y)}\ \frac{T_1(\cdot)}{1-T_1(\cdot)}\ \frac{y_1}{\varepsilon_2} \ , \tag{136}$$

and

$$\bar{\beta}^0_{I,2}(y) = -\frac{\sigma^{12,0}(\cdot)}{\bar{H}(y)}\ \frac{T_2(\cdot)}{1-T_2(\cdot)}\ \frac{y_2}{\varepsilon_1} \ . \tag{137}$$

69. The Lemma follows from the analysis in Appendix K, see Lemma 4 and Corollary 2.

70. If solution to the choice problem in (5) is unique and satisfies first order conditions then $\bar{H}(y) > 0$.

Moreover, $\bar{\mathcal{R}}_b(y) = (\bar{\mu}_b^0(y) + \bar{\beta}_{E,b}^0(y))f(y)$ *where*

$$\bar{\mu}_b^0(y) + \bar{\beta}_{E,b}^0(y) = \sigma^{12,0}(\bar{\omega}(y))\Big(1 - L_E(\bar{\omega}(y))\Big) \tag{138}$$

and

$$\begin{aligned} L_E(\bar{\omega}(y)) \;\; = \;\; & \chi_{12,1}^0(\bar{\omega}(y))\big(T(y) - T(y_{s1}^0(\bar{\omega_1}(y), 0))\big) \\ & + \chi_{12,2}^0(\bar{\omega}(y))\big(T(y) - T(0, y_{s2}^0(0, \bar{\omega_2}(y)))\big) \; . \end{aligned}$$

Inputs for evaluating tax reforms Based on Lemma 6, and under a piece-wise linear tax system, inputs for the evaluation of tax reforms simplify to

$$\begin{aligned} \bar{\mathcal{R}}_1(y_1, y_2) &= -f(y_1, y_2)\; \varepsilon_1\; y_1\; \frac{T_1(y_1, y_2)}{1 - T_1(y_1, y_2)} \\ \bar{\mathcal{R}}_2(y_1, y_2) &= -f(y_1, y_2)\; \varepsilon_2\; y_2\; \frac{T_2(y_1, y_2)}{1 - T_2(y_1, y_2)} \\ \bar{\mathcal{R}}_b(y_1, y_2) &= f(y_1, y_2)\big(1 - L_E(y_1, y_2)\big). \end{aligned} \tag{139}$$

where $f(y_1, y_2) = \sigma^{12,0} f(y)$ denotes the observed mass of couples with primary (secondary) earnings y_1 (y_2). We will use these inputs when analyzing inefficiencies and welfare consequences of different types of tax reforms.

L.3 Lowering marginal tax rates on secondary earnings

In Section 6.1 of the main text, we analyze the effects of tax reforms that reduce marginal tax rates for secondary earnings using conditional revenue functions. In this section of the appendix, we provide more detailed explanations on how empirical revenue functions are constructed (Section L.3.1). We also show supplementary results on the empirical ingredients of these conditional revenue functions (Section L.3.2), robustness tests regarding behavioral responses at the extensive and intensive margin (Section L.3.3), and results for alternative choices of the conditioning brackets (Section L.3.4). We also provide additional insights on the welfare effects of lowering marginal tax rates at the bottom (Section L.3.5).

L.3.1 Revenue and welfare functions

To identify inefficiencies in the status quo tax system, we follow equation (20) and estimate (un)conditional revenue functions for primary and secondary earners. *Conditional* revenue functions describe the implications of reforms that increase marginal tax rates on the earnings of one spouse y_i for all couples where the other spouse has income in a specific income bracket, i.e., $y_{-i} \in B_{-i} = [\underline{b}_{-i}, \overline{b}_{-i}]$.[71] Drawing on the inputs from equation (139), and focusing on dual-earner couples, the empirical conditional revenue function for secondary earnings reads

$$\begin{aligned}\mathbf{R}_2(y_2 \mid B_1) &= \int_{y_1 \in B_1} \mathbf{R}_2(y_2 \mid y_1)\, dy_1 \;= \int_{y_1 \in B_1} \quad \bar{\mathcal{R}}_2(y_1, y_2) + \int_{y_2}^{\bar{y}_2} \bar{\mathcal{R}}_b(y_1, s_2)\, ds_2 \;\; dy_1 \\ &= s(B_1) \left(\mathbb{E}_{y_1 \in B_1}\Big[- f(y_2|y_1)\, \varepsilon_2\, y_2\, \frac{T_2(y_1, y_2)}{1 - T_2(y_1, y_2)} \Big] + \mathbb{E}_{y_1 \in B_1}\Big[\big(1 - F(y_2|y_1)\big)\big(1 - L_E\big) \Big] \right)\end{aligned} \tag{140}$$

where $s_1(B_1) = F_{y_1}(\overline{b}_1) - F_{y_1}(\underline{b}_1)$ is the mass of couples with $y_1 \in B_1 = [\underline{b}_1, \overline{b}_1]$. In contrast, *unconditional* revenue functions describe the implications of reforms that modify the marginal tax rate for the earnings of one spouse independent from the earnings of the other spouse. The secondary earnings unconditional revenue function reads

$$\begin{aligned}\mathbf{R}_2(y_2) &= \int_{y_1} \mathbf{R}_2(y_2 \mid y_1)\, dy_1 \\ &= -f(y_2)\, \varepsilon_2\, y_2\, \mathbb{E}_{y_1}\Big[\frac{T_2(y_1, y_2)}{1 - T_2(y_1, y_2)} \Big] + \big(1 - F(y_2)\big)\big(1 - L_E\big)\,.\end{aligned} \tag{141}$$

For the empirical application, we assume that the conditional distribution of secondary earnings does not vary within the specific bracket of primary earnings that we are looking at. This assumption is necessary, since we require enough data points to estimate the secondary earner income distribution. We also assume that the extensive margin responses do not vary across the conditioning earnings variable. Consequently, the empirically estimated revenue function simplifies to

71. We condition on an income bracket B_{-i}, because data constraints make conditioning on exactly one income level y_{-i} infeasible. In particular, the estimation of conditional probability density functions becomes very noisy as soon as the conditioning variable is limited to one exact dollar value. For additional details on the estimation, see Appendix L.2.

$$\mathbf{R}_2(y_2 \mid \underline{y}_1^Q \leq y_{1s} < \overline{y}_1^Q) = \underbrace{s_1^Q}_{\text{Share}} \Bigg(\underbrace{- \mathbb{1}_2^d \, f(y_2 \mid \underline{y}_1^Q \leq y_{1s} < \overline{y}_1^Q)}_{\text{Intensive}} + \underbrace{\Big(1 - F(y_2 \mid \underline{y}_1^Q \leq y_{1s} < \overline{y}_1^Q)\Big)}_{\text{Cond. CDF}} \underbrace{\Big(1 - L_E\Big)}_{\text{Extensive}} \Bigg). \tag{142}$$

$$\mathbb{1}_2^d = \mathbb{E}_{y_1}\Big[\varepsilon_2 \, y_2 \, \frac{T_2(y_1, y_2)}{1 - T_2(y_1, y_2)} \mid \underline{y}_1^Q \leq y_{1s} < \overline{y}_1^Q\Big]$$

where $\underline{y}_1^Q$ and $\overline{y}_1^Q$ indicate lower and upper thresholds for the respective primary earner income range Q, upon which the revenue functions are conditioned.

The equation provides an intuitive understanding of the mechanics behind the conditional revenue function. The term s_1^Q reflects the share of dual earner couples whose primary earnings fall in the respective income range, and are thus affected by a tax reform. The first part of the equation captures the change in tax revenue through behavioral responses at the intensive margin. The intensive margin behavioral response affects all secondary earners at the income level y_2 with primary earnings in the respective income range. $\mathbb{1}_2^d$ describes the reaction of these secondary earners, and is based on an average at every secondary earnings level y_2, because marginal tax rates faced by secondary earners, as well as elasticities, might vary across primary earnings in the bracket. The second part of the equation captures the mechanical change in tax revenue. This mechanical change in tax revenue is downscaled by the behavioral response at the extensive margin captured by the extensive margin parameter L_E.

Unconditional revenue functions We also compare *conditional* revenue functions to the respective *unconditional* revenue function. The unconditional revenue function illustrates the potential for Pareto improvements through tax rate changes that are independent from the earnings of the other spouse. For secondary earners in dual earner couples, this unconditional revenue function reads

$$\begin{aligned}\mathbf{R}_2(y_2) &= \int_{y_1 \in B_1} \mathbf{R}_2(y_2 \mid y_1)\, dy_1 \\ &= \underbrace{-f(y_2)\, \mathbb{E}_{y_1}\Big[\varepsilon_2\, y_2\, \frac{T_2(y_1, y_2)}{1 - T_2(y_1, y_2)}\Big]}_{\text{Intensive Margin}} + \underbrace{\big(1 - F(y_2)\big)}_{\text{Mechanical Effect}} \quad \underbrace{\big(1 - L_E\big)}_{\text{Extensive Margin}} \quad . \end{aligned} \tag{143}$$

Together, the value and shape of the empirical conditional and unconditional revenue functions characterize the potential for Pareto improving tax reforms for secondary earners with a spouse in specific income ranges.

A comparison between conditional and unconditional revenue functions reveals whether inefficiencies in the tax code can be cured by changing marginal tax rates for *all* secondary earners or whether curing inefficiencies requires the joint consideration of secondary and primary earnings.

Conditional welfare functions Next to conditional revenue functions, we also estimate conditional welfare functions, which provide an estimate of the welfare consequences of tax reforms. The conditional welfare function reads

$$\mathbf{W}_2(y_2 \mid y_1) \quad = \quad \mathcal{R}_2(y_2 \mid y_1) + \big(1 - F(y_2 \mid y_1)\big)\big(1 - \mathcal{G}(y_2 \mid y_1)\big) f_1(y_1) \tag{144}$$

where $\mathcal{G}(y_2 \mid y_1)$ is the average welfare weight for married couples with primary earnings y_1 and secondary earnings above y_2. As for the revenue functions, we use primary earnings income brackets in the empirical application. Information on the specification of welfare weights is provided in Table L5.

L.3.2 Ingredients

The estimation of revenue functions requires assumptions about labor supply elasticities and conditioning brackets for the reform. In addition, we need estimates on the (conditional) income distributions and intensive margin response terms.

Table L2 display three different scenarios that we use throughout the paper and that are informed by the literature listed in the table note. Table L3 presents the upper thresholds of the income deciles for the primary and secondary earner income distributions that are used as conditioning brackets.

We estimate income distributions by means of a kernel density estimation using a Gaussian kernel and 20,000 grid points. The income grid ranges from zero to one million dollars.

The intensive margin response terms in equation (142) and (143) demand knowledge of a conditional average. For instance, the secondary earner conditional revenue function contains an average of the intensive margin response at every secondary earnings income level conditional on primary earnings in a specific decile. We estimate these averages by means of fitting a local polynomial regression of degree zero (local mean smoothing) with a Gaussian kernel as a weighting function.

Figures L6 - L15 show the empirical ingredients to compute the conditional revenue functions for secondary earners, in particular marginal tax rates, conditional income distributions, and behavioral responses to taxation. Figures L16-L25 show the respective ingredients for the conditional revenue functions for primary earners.

Table L2: Assumptions about labor supply elasticities

	Primary earner	Secondary earner
Low elasticity scenario	0.15	0.35
Baseline elasticity scenario	0.25	0.75
High elasticity scenario	0.5	1.5

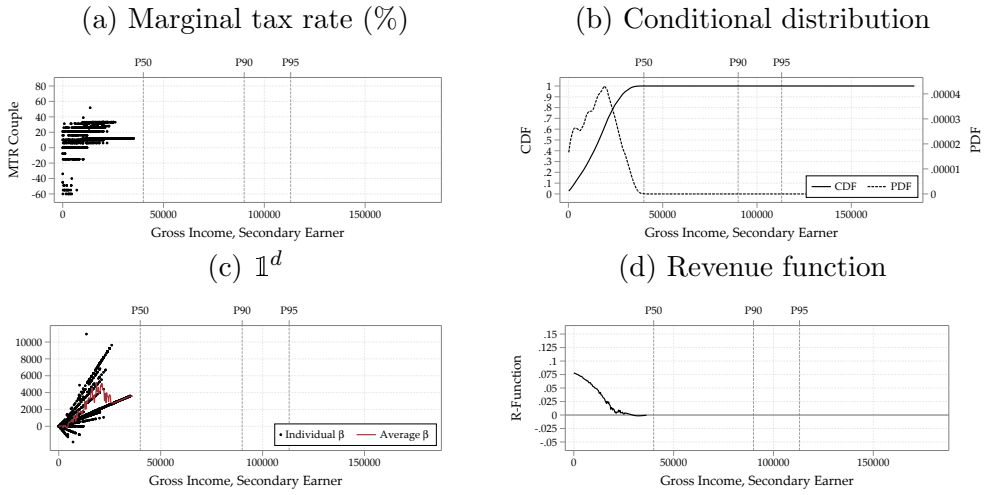
Note: This table displays our assumptions about the labor supply elasticities for primary and secondary earners in married couples. Assumptions are guided by the range of estimates found in the literature, e.g. Blundell and Macurdy (1999), Blau and Kahn (2007), Eissa and Hoynes (2004), LaLumia (2008), Saez, Slemrod, and Giertz (2012), Bargain, Orsini, and Peichl (2014), and Neisser (2021).

Table L3: Income deciles (2019)

Decile	Primary earnings	Secondary earnings
Q1	35000	10001
Q2	46002	19000
Q3	55604	25015
Q4	65228	32005
Q5	77056	40000
Q6	91032	47300
Q7	108966	55009
Q8	135030	67785
Q9	184155	90001
Q10	18009756	1358632

Note: This table shows the upper thresholds of the deciles for the primary and secondary earner income distribution among dual earner couples as of 2019. All estimates are based on tax units with non-negative gross income in which both spouses are between 25 and 55 years old. Our pre-tax gross income variable of interest contains wage income, farm income, business income, income from dividends, income from interest, income from rent, and retirement income.
Source: Own calculations based on CPS-ASEC (2019)

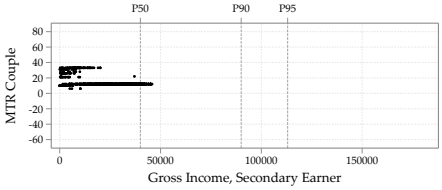
Figure L6: Conditional revenue functions, sec. earners, PE Q1



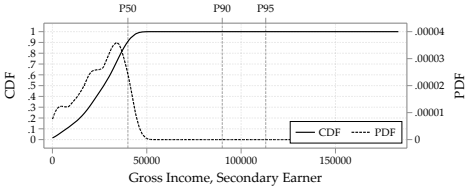
Note: This figure shows ingredients of equation (142) – marginal tax rate (Panel a), conditional income distribution (Panel b), intensive margin response $\mathbb{1}^d$ (Panel c), and conditional revenue function (Panel d) – among dual earner couples as of 2019 for primary earnings in decile Q1. All estimates are based on tax units with non-negative gross income in which spouses are of age 25-55. Gross income contains wage, farm, retirement, and business income, as well as income from dividends, interest, and rent. We take behavioral responses with $\epsilon_1 = 0.25$ and $L_E = 0.2$ into account.
Source: Own calculations based on CPS-ASEC (2019)

Figure L7: Conditional revenue functions, sec. earners, PE Q2

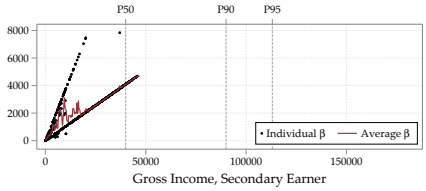
(a) Marginal tax rate (%)



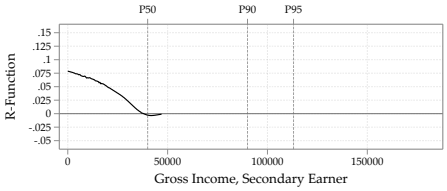
(b) Conditional distribution



(c) $\mathbb{1}^d$



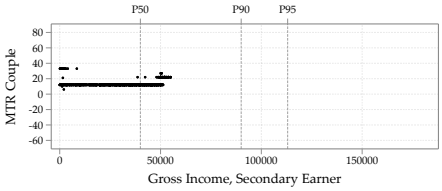
(d) Revenue function



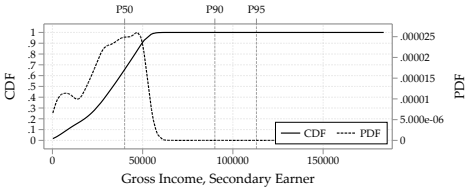
Note: This figure shows ingredients of equation (142) – marginal tax rate (Panel a), conditional income distribution (Panel b), intensive margin response $\mathbb{1}^d$ (Panel c), and conditional revenue function (Panel d) – among dual earner couples as of 2019 for primary earnings in decile Q2. All estimates are based on tax units with non-negative gross income in which spouses are of age 25-55. Gross income contains wage, farm, retirement, and business income, as well as income from dividends, interest, and rent. We take behavioral responses with $\epsilon_2 = 0.75$ and $L_E = 0.2$ into account.
Source: Own calculations based on CPS-ASEC (2019)

Figure L8: Conditional revenue functions, sec. earners, PE Q3

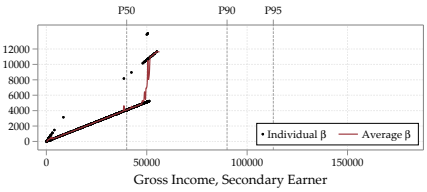
(a) Marginal tax rate (%)



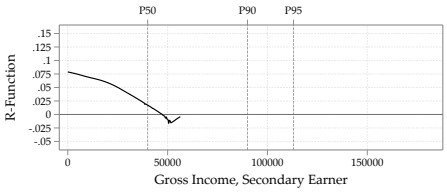
(b) Conditional distribution



(c) $\mathbb{1}^d$



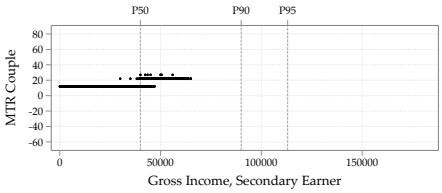
(d) Revenue function



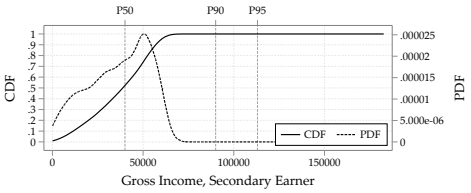
Note: This figure shows ingredients of equation (142) – marginal tax rate (Panel a), conditional income distribution (Panel b), intensive margin response $\mathbb{1}^d$ (Panel c), and conditional revenue function (Panel d) – among dual earner couples as of 2019 for primary earnings in decile Q3. All estimates are based on tax units with non-negative gross income in which spouses are of age 25-55. Gross income contains wage, farm, retirement, and business income, as well as income from dividends, interest, and rent. We take behavioral responses with $\epsilon_2 = 0.75$ and $L_E = 0.2$ into account.
Source: Own calculations based on CPS-ASEC (2019)

Figure L9: Conditional revenue functions, sec. earners, PE Q4

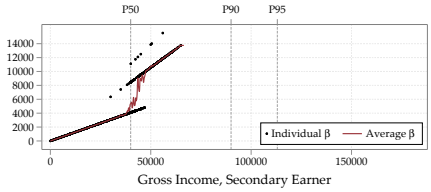
(a) Marginal tax rate (%)



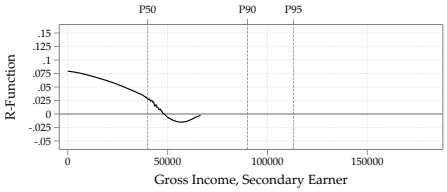
(b) Conditional distribution



(c) $\mathbb{1}^d$



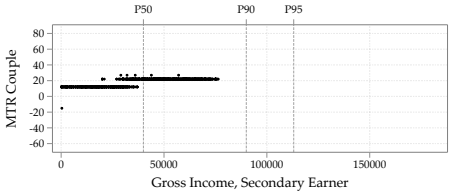
(d) Revenue function



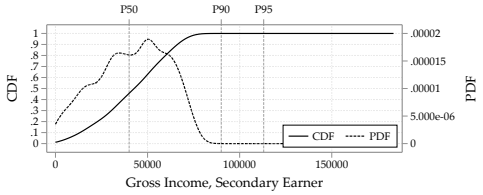
Note: This figure shows ingredients of equation (142) – marginal tax rate (Panel a), conditional income distribution (Panel b), intensive margin response $\mathbb{1}^d$ (Panel c), and conditional revenue function (Panel d) – among dual earner couples as of 2019 for primary earnings in decile Q4. All estimates are based on tax units with non-negative gross income in which spouses are of age 25-55. Gross income contains wage, farm, retirement, and business income, as well as income from dividends, interest, and rent. We take behavioral responses with $\epsilon_2 = 0.75$ and $L_E = 0.2$ into account.
Source: Own calculations based on CPS-ASEC (2019)

Figure L10: Conditional revenue functions, Q5 PE, Ingredients

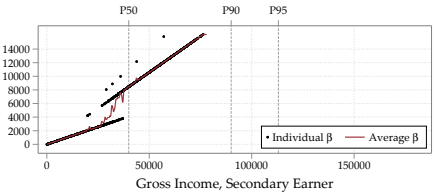
(a) Marginal tax rate (%)



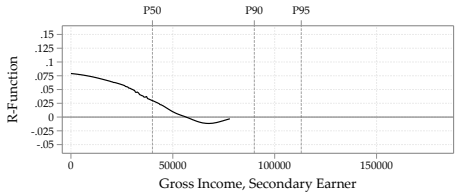
(b) Conditional distribution



(c) $\mathbb{1}^d$

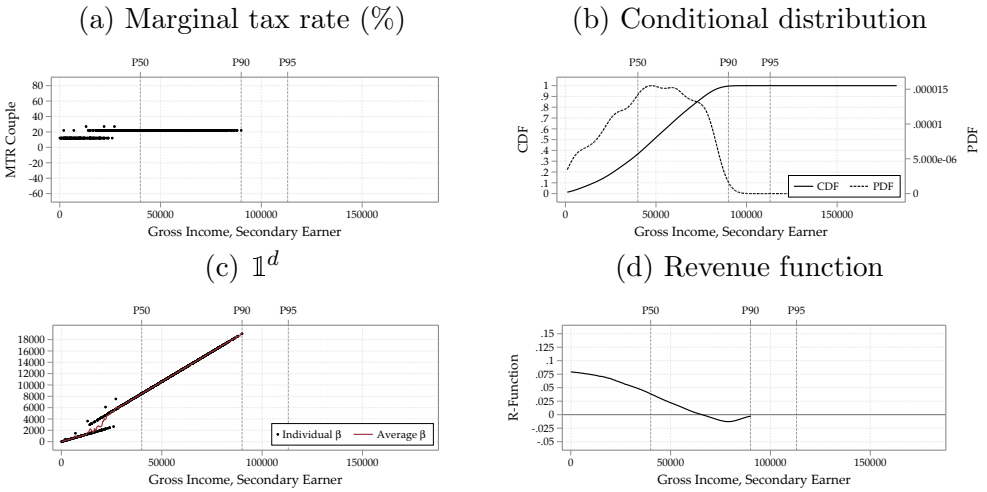


(d) Revenue function



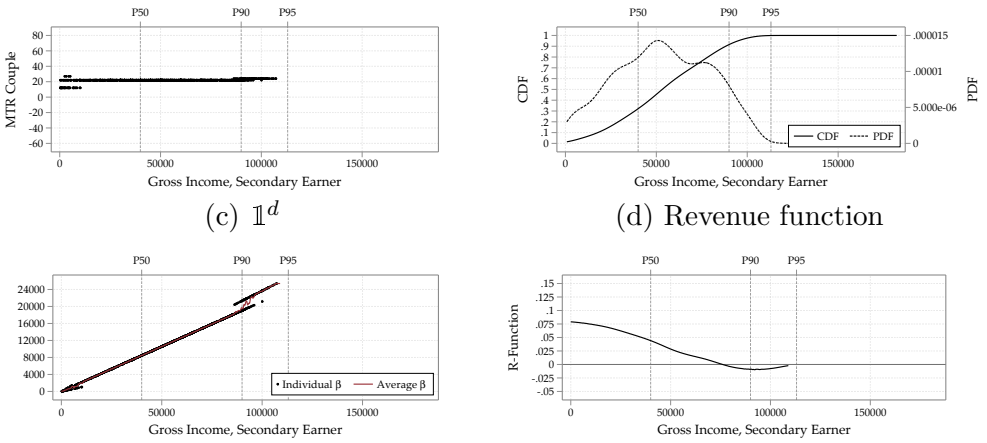
Note: This figure shows ingredients of equation (142) – marginal tax rate (Panel a), conditional income distribution (Panel b), intensive margin response $\mathbb{1}^d$ (Panel c), and conditional revenue function (Panel d) – among dual earner couples as of 2019 for primary earnings in decile Q5. All estimates are based on tax units with non-negative gross income in which spouses are of age 25-55. Gross income contains wage, farm, retirement, and business income, as well as income from dividends, interest, and rent. We take behavioral responses with $\epsilon_2 = 0.75$ and $L_E = 0.2$ into account.
Source: Own calculations based on CPS-ASEC (2019)

Figure L11: Conditional revenue functions, sec. earners, PE Q6



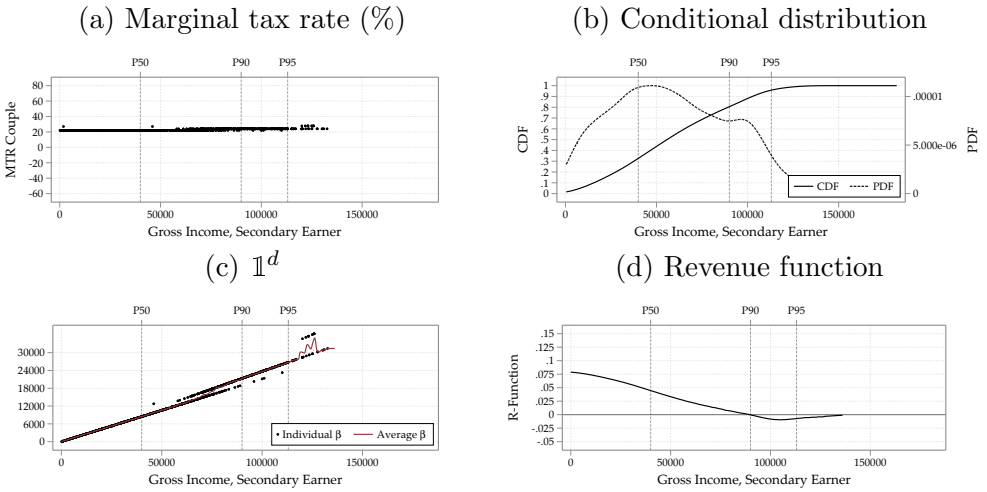
Note: This figure shows ingredients of equation (142) – marginal tax rate (Panel a), conditional income distribution (Panel b), intensive margin response $\mathbb{1}^d$ (Panel c), and conditional revenue function (Panel d) – among dual earner couples as of 2019 for primary earnings in decile Q6. All estimates are based on tax units with non-negative gross income in which spouses are of age 25-55. Gross income contains wage, farm, retirement, and business income, as well as income from dividends, interest, and rent. We take behavioral responses with $\epsilon_2 = 0.75$ and $L_E = 0.2$ into account.
Source: Own calculations based on CPS-ASEC (2019)

Figure L12: Conditional revenue functions, sec. earners, PE Q7



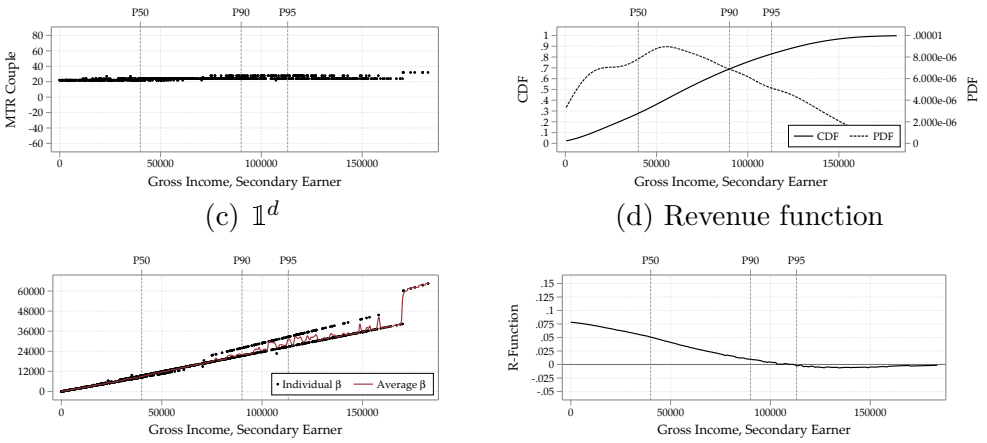
Note: This figure shows ingredients of equation (142) – marginal tax rate (Panel a), conditional income distribution (Panel b), intensive margin response $\mathbb{1}^d$ (Panel c), and conditional revenue function (Panel d) – among dual earner couples as of 2019 for primary earnings in decile Q7. All estimates are based on tax units with non-negative gross income in which spouses are of age 25-55. Gross income contains wage, farm, retirement, and business income, as well as income from dividends, interest, and rent. We take behavioral responses with $\epsilon_2 = 0.75$ and $L_E = 0.2$ into account.
Source: Own calculations based on CPS-ASEC (2019)

Figure L13: Conditional revenue functions, sec. earners, PE Q8



Note: This figure shows ingredients of equation (142) – marginal tax rate (Panel a), conditional income distribution (Panel b), intensive margin response $\mathbb{1}^d$ (Panel c), and conditional revenue function (Panel d) – among dual earner couples as of 2019 for primary earnings in decile Q8. All estimates are based on tax units with non-negative gross income in which spouses are of age 25-55. Gross income contains wage, farm, retirement, and business income, as well as income from dividends, interest, and rent. We take behavioral responses with $\epsilon_2 = 0.75$ and $L_E = 0.2$ into account.
Source: Own calculations based on CPS-ASEC (2019)

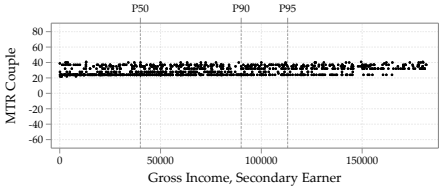
Figure L14: Conditional revenue functions, sec. earners, PE Q9



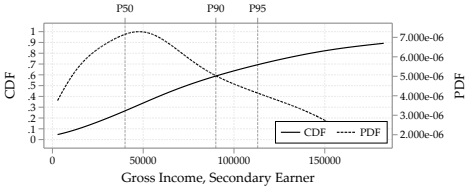
Note: This figure shows ingredients of equation (142) – marginal tax rate (Panel a), conditional income distribution (Panel b), intensive margin response $\mathbb{1}^d$ (Panel c), and conditional revenue function (Panel d) – among dual earner couples as of 2019 for primary earnings in decile Q9. All estimates are based on tax units with non-negative gross income in which spouses are of age 25-55. Gross income contains wage, farm, retirement, and business income, as well as income from dividends, interest, and rent. We take behavioral responses with $\epsilon_2 = 0.75$ and $L_E = 0.2$ into account.
Source: Own calculations based on CPS-ASEC (2019)

Figure L15: Conditional revenue functions, sec. earners, PE Q10

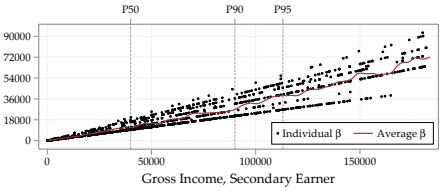
(a) Marginal tax rate (%)



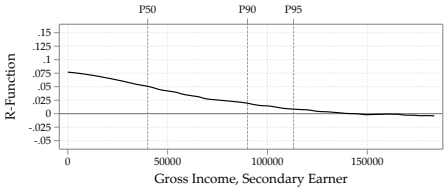
(b) Conditional distribution



(c) $\mathbb{1}^d$



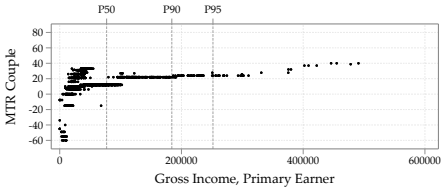
(d) Revenue function



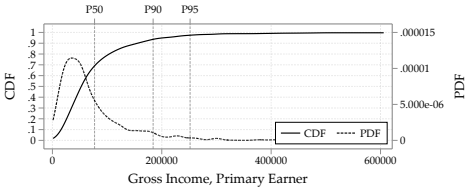
Note: This figure shows ingredients of equation (142) – marginal tax rate (Panel a), conditional income distribution (Panel b), intensive margin response $\mathbb{1}^d$ (Panel c), and conditional revenue function (Panel d) – among dual earner couples as of 2019 for primary earnings in decile Q10. All estimates are based on tax units with non-negative gross income in which spouses are of age 25-55. Gross income contains wage, farm, retirement, and business income, as well as income from dividends, interest, and rent. We take behavioral responses with $\epsilon_2 = 0.75$ and $L_E = 0.2$ into account.
Source: Own calculations based on CPS-ASEC (2019)

Figure L16: Conditional revenue functions, prim. earners, SE Q1

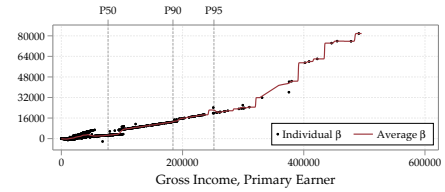
(a) Marginal tax rate (%)



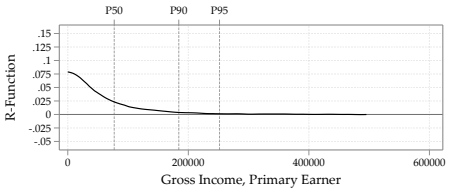
(b) Conditional distribution



(c) $\mathbb{1}^d$



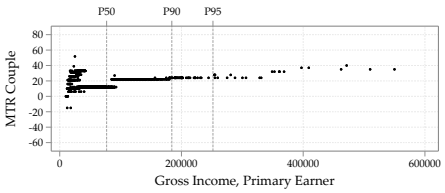
(d) Revenue function



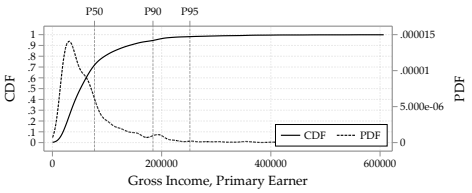
Note: This figure shows ingredients of equation (142) – marginal tax rate (Panel a), conditional income distribution (Panel b), intensive margin response $\mathbb{1}^d$ (Panel c), and conditional revenue function (Panel d) – among dual earner couples as of 2019 for secondary earnings in decile Q1. All estimates are based on tax units with non-negative gross income in which spouses are of age 25-55. Gross income contains wage, farm, retirement, and business income, as well as income from dividends, interest, and rent. We take behavioral responses with $\epsilon_1 = 0.25$ and $L_E = 0.2$ into account.
Source: Own calculations based on CPS-ASEC (2019)

Figure L17: Conditional revenue functions, prim. earners, SE Q2

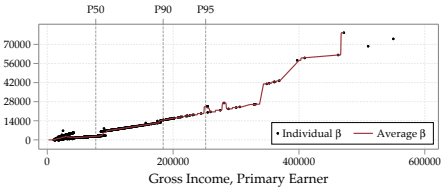
(a) Marginal tax rate (%)



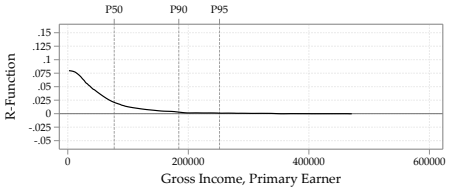
(b) Conditional distribution



(c) $\mathbb{1}^d$



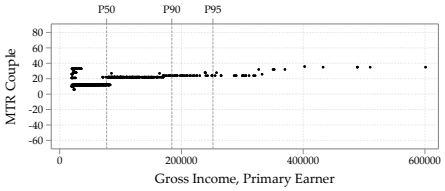
(d) Revenue function



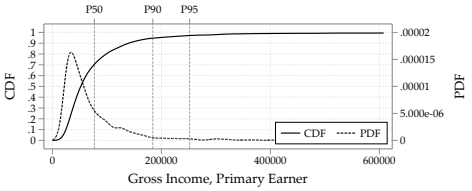
Note: This figure shows ingredients of equation (142) – marginal tax rate (Panel a), conditional income distribution (Panel b), intensive margin response $\mathbb{1}^d$ (Panel c), and conditional revenue function (Panel d) – among dual earner couples as of 2019 for secondary earnings in decile Q2. All estimates are based on tax units with non-negative gross income in which spouses are of age 25-55. Gross income contains wage, farm, retirement, and business income, as well as income from dividends, interest, and rent. We take behavioral responses with $\epsilon_1 = 0.25$ and $L_E = 0.2$ into account.
Source: Own calculations based on CPS-ASEC (2019)

Figure L18: Conditional revenue functions, prim. earners, SE Q3

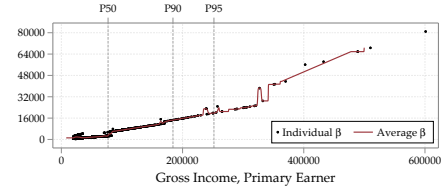
(a) Marginal tax rate (%)



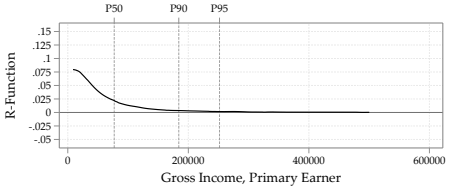
(b) Conditional distribution



(c) $\mathbb{1}^d$



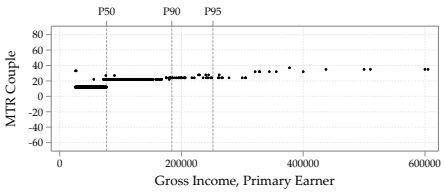
(d) Revenue function



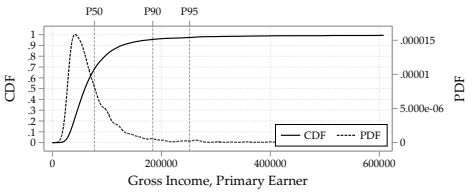
Note: This figure shows ingredients of equation (142) – marginal tax rate (Panel a), conditional income distribution (Panel b), intensive margin response $\mathbb{1}^d$ (Panel c), and conditional revenue function (Panel d) – among dual earner couples as of 2019 for secondary earnings in decile Q3. All estimates are based on tax units with non-negative gross income in which spouses are of age 25-55. Gross income contains wage, farm, retirement, and business income, as well as income from dividends, interest, and rent. We take behavioral responses with $\epsilon_1 = 0.25$ and $L_E = 0.2$ into account.
Source: Own calculations based on CPS-ASEC (2019)

Figure L19: Conditional revenue functions, prim. earners, SE Q4

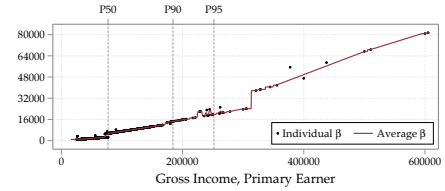
(a) Marginal tax rate (%)



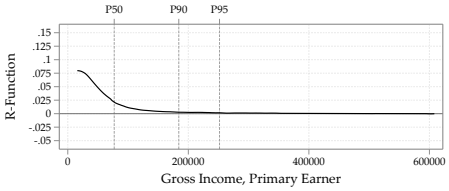
(b) Conditional distribution



(c) $\mathbb{1}^d$



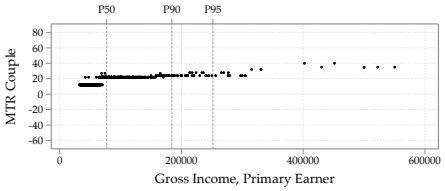
(d) Revenue function



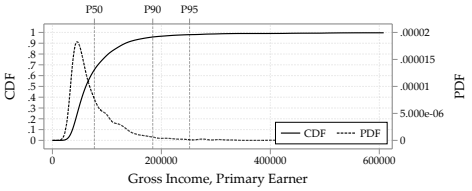
Note: This figure shows ingredients of equation (142) – marginal tax rate (Panel a), conditional income distribution (Panel b), intensive margin response $\mathbb{1}^d$ (Panel c), and conditional revenue function (Panel d) – among dual earner couples as of 2019 for secondary earnings in decile Q4. All estimates are based on tax units with non-negative gross income in which spouses are of age 25-55. Gross income contains wage, farm, retirement, and business income, as well as income from dividends, interest, and rent. We take behavioral responses with $\epsilon_1 = 0.25$ and $L_E = 0.2$ into account.
Source: Own calculations based on CPS-ASEC (2019)

Figure L20: Conditional revenue functions, prim. earners, SE Q5

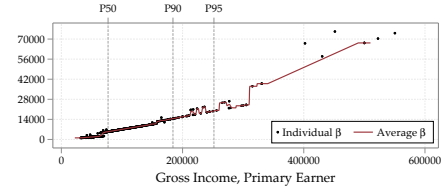
(a) Marginal tax rate (%)



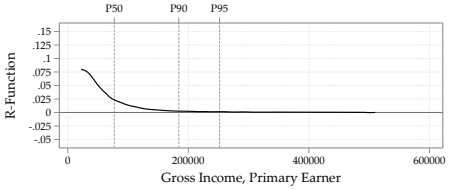
(b) Conditional distribution



(c) $\mathbb{1}^d$



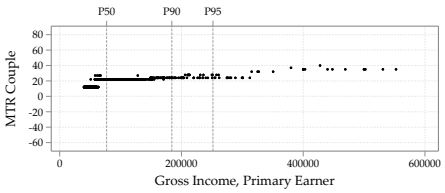
(d) Revenue function



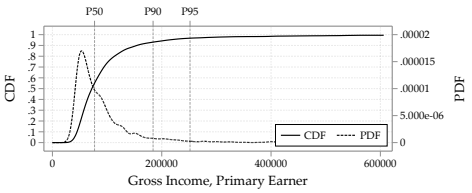
Note: This figure shows ingredients of equation (142) – marginal tax rate (Panel a), conditional income distribution (Panel b), intensive margin response $\mathbb{1}^d$ (Panel c), and conditional revenue function (Panel d) – among dual earner couples as of 2019 for secondary earnings in decile Q5. All estimates are based on tax units with non-negative gross income in which spouses are of age 25-55. Gross income contains wage, farm, retirement, and business income, as well as income from dividends, interest, and rent. We take behavioral responses with $\epsilon_1 = 0.25$ and $L_E = 0.2$ into account.
Source: Own calculations based on CPS-ASEC (2019)

Figure L21: Conditional revenue functions, prim. earners, SE Q6

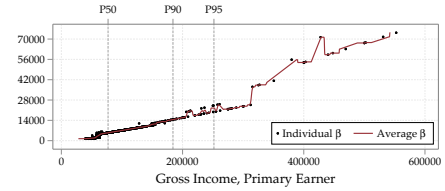
(a) Marginal tax rate (%)



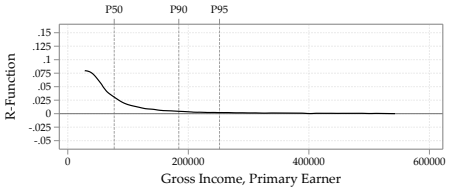
(b) Conditional distribution



(c) $\mathbb{1}^d$



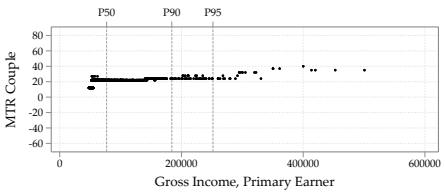
(d) Revenue function



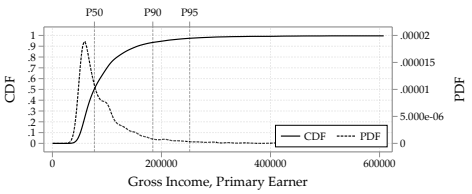
Note: This figure shows ingredients of equation (142) – marginal tax rate (Panel a), conditional income distribution (Panel b), intensive margin response $\mathbb{1}^d$ (Panel c), and conditional revenue function (Panel d) – among dual earner couples as of 2019 for secondary earnings in decile Q6. All estimates are based on tax units with non-negative gross income in which spouses are of age 25-55. Gross income contains wage, farm, retirement, and business income, as well as income from dividends, interest, and rent. We take behavioral responses with $\epsilon_1 = 0.25$ and $L_E = 0.2$ into account.
Source: Own calculations based on CPS-ASEC (2019)

Figure L22: Conditional revenue functions, prim. earners, SE Q7

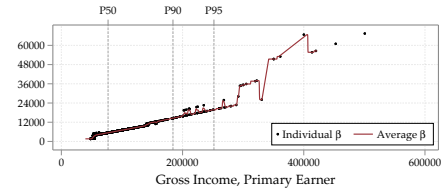
(a) Marginal tax rate (%)



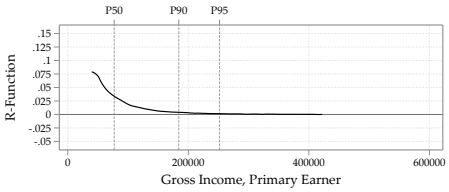
(b) Conditional distribution



(c) $\mathbb{1}^d$



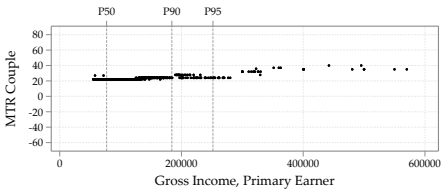
(d) Revenue function



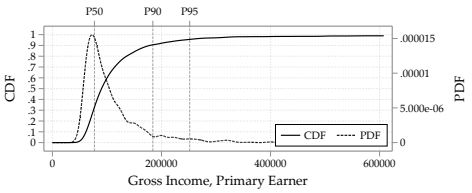
Note: This figure shows ingredients of equation (142) – marginal tax rate (Panel a), conditional income distribution (Panel b), intensive margin response $\mathbb{1}^d$ (Panel c), and conditional revenue function (Panel d) – among dual earner couples as of 2019 for secondary earnings in decile Q7. All estimates are based on tax units with non-negative gross income in which spouses are of age 25-55. Gross income contains wage, farm, retirement, and business income, as well as income from dividends, interest, and rent. We take behavioral responses with $\epsilon_1 = 0.25$ and $L_E = 0.2$ into account.
Source: Own calculations based on CPS-ASEC (2019)

Figure L23: Conditional revenue functions, prim. earners, SE Q8

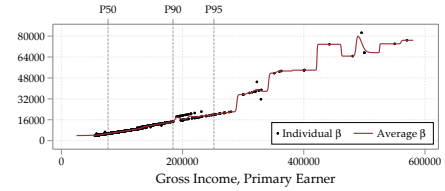
(a) Marginal tax rate (%)



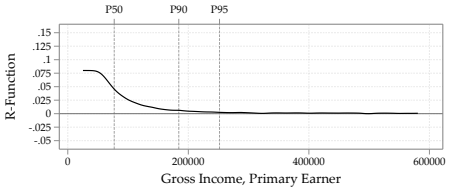
(b) Conditional distribution



(c) $\mathbb{1}^d$



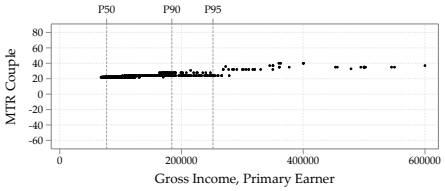
(d) Revenue function



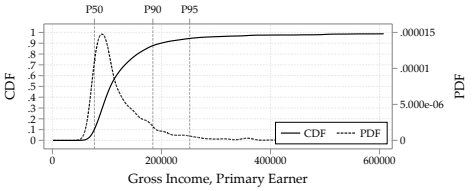
Note: This figure shows ingredients of equation (142) – marginal tax rate (Panel a), conditional income distribution (Panel b), intensive margin response $\mathbb{1}^d$ (Panel c), and conditional revenue function (Panel d) – among dual earner couples as of 2019 for secondary earnings in decile Q8. All estimates are based on tax units with non-negative gross income in which spouses are of age 25-55. Gross income contains wage, farm, retirement, and business income, as well as income from dividends, interest, and rent. We take behavioral responses with $\epsilon_1 = 0.25$ and $L_E = 0.2$ into account.
Source: Own calculations based on CPS-ASEC (2019)

Figure L24: Conditional revenue functions, prim. earners, SE Q9

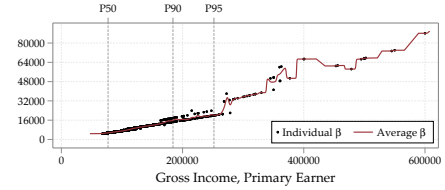
(a) Marginal tax rate (%)



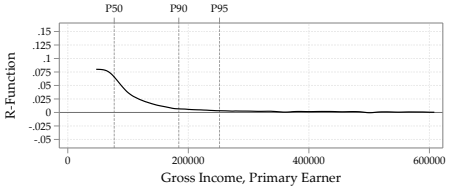
(b) Conditional distribution



(c) $\mathbb{1}^d$

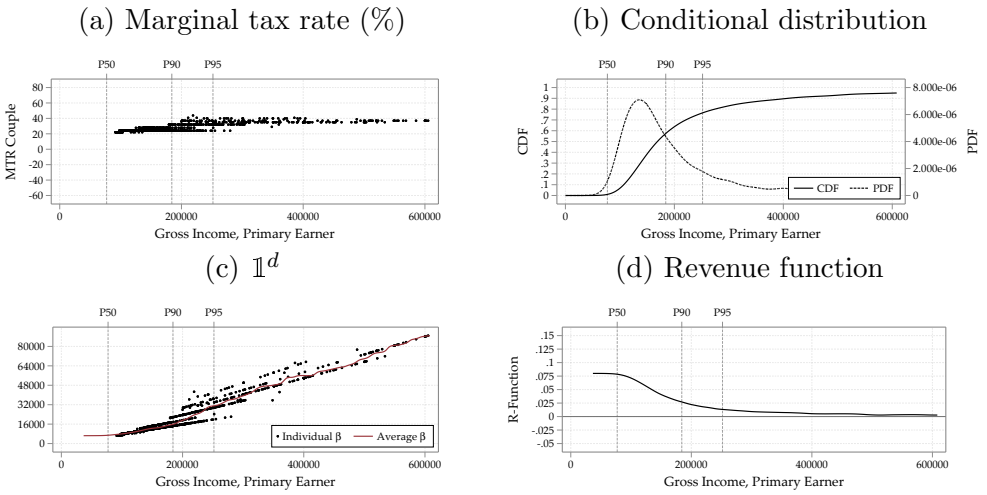


(d) Revenue function



Note: This figure shows ingredients of equation (142) – marginal tax rate (Panel a), conditional income distribution (Panel b), intensive margin response $\mathbb{1}^d$ (Panel c), and conditional revenue function (Panel d) – among dual earner couples as of 2019 for secondary earnings in decile Q9. All estimates are based on tax units with non-negative gross income in which spouses are of age 25-55. Gross income contains wage, farm, retirement, and business income, as well as income from dividends, interest, and rent. We take behavioral responses with $\epsilon_1 = 0.25$ and $L_E = 0.2$ into account.
Source: Own calculations based on CPS-ASEC (2019)

Figure L25: Conditional revenue functions, prim. earners, SE Q10



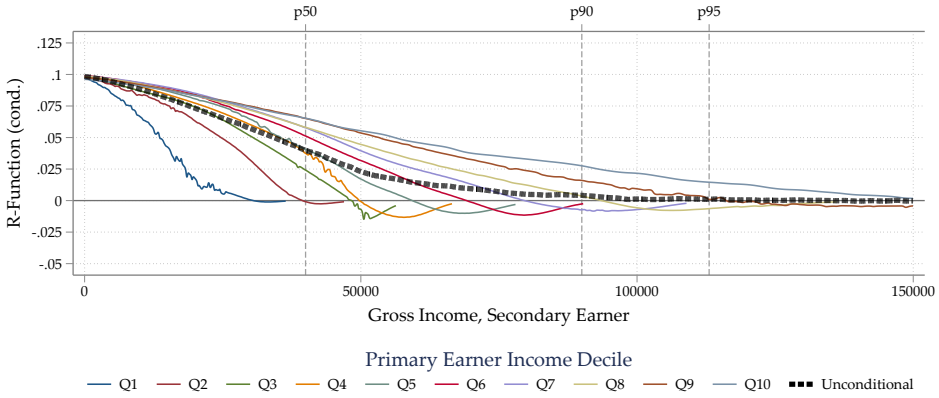
Note: This figure shows ingredients of equation (142) – marginal tax rate (Panel a), conditional income distribution (Panel b), intensive margin response $\mathbb{1}^d$ (Panel c), and conditional revenue function (Panel d) – among dual earner couples as of 2019 for secondary earnings in decile Q10. All estimates are based on tax units with non-negative gross income in which spouses are of age 25-55. Gross income contains wage, farm, retirement, and business income, as well as income from dividends, interest, and rent. We take behavioral responses with $\epsilon_1 = 0.25$ and $L_E = 0.2$ into account.
Source: Own calculations based on CPS-ASEC (2019)

L.3.3 Robustness

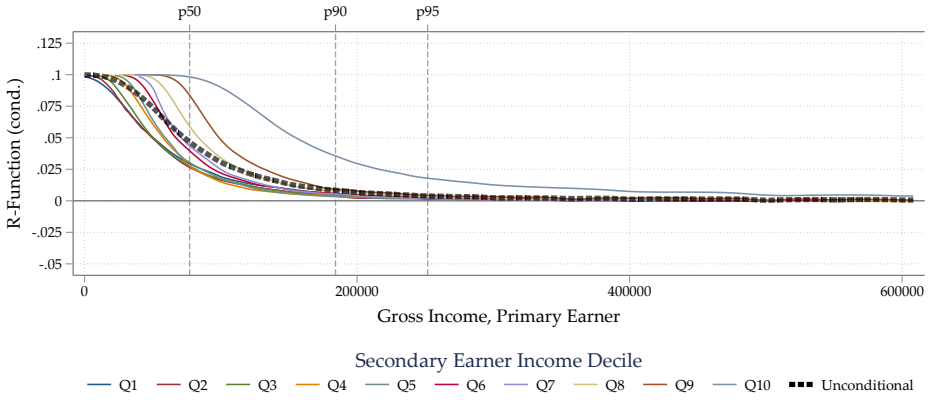
In the baseline, conditional revenue functions are computed under the assumption of extensive margin and moderate intensive margin responses. We provide robustness tests regarding the exclusion of margin responses (Figure L26) and for the use of low (high) elasticity scenarios from Table L2 in Figure L27 (L28).

Figure L26: Cond. revenue functions, deciles (2019), no extensive margin

(a) Secondary earners



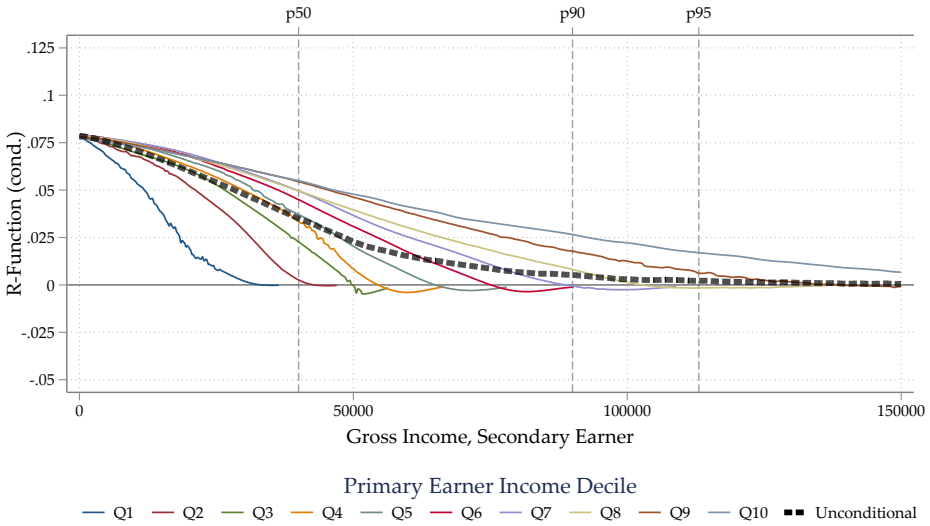
(b) Primary earners



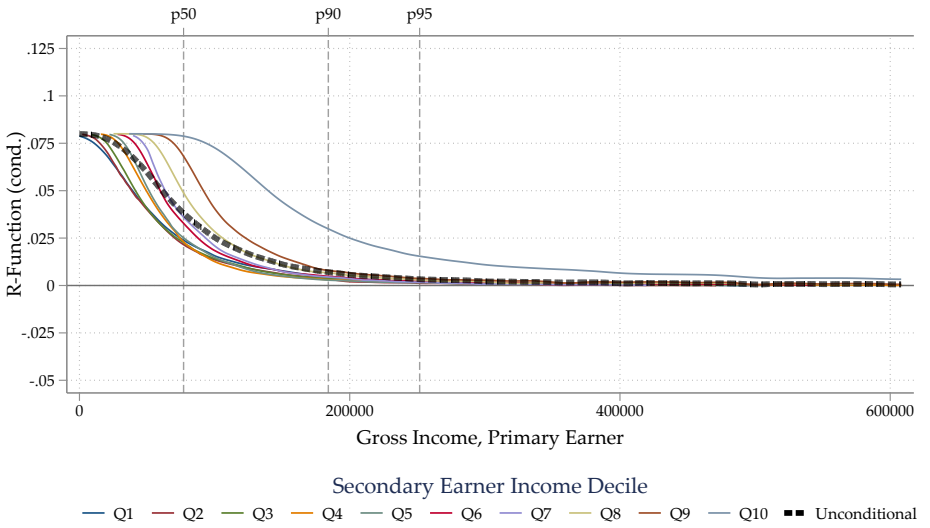
Note: This figure shows conditional revenue functions for secondary (primary) earners in married dual earner couples conditional on primary (secondary) earnings income deciles in Panel a (b) as of 2019. All estimates are based on tax units with non-negative gross income in which spouses are of age 25-55. Gross income contains wage, farm, retirement, and business income, as well as income from dividends, interest, and rent. We take behavioral responses at the intensive margin into account. We assume intensive margin elasticities of 0.25 (0.75) for primary (secondary) earners. The figure also displays modified unconditional revenue functions for secondary and primary earners where unconditional revenue functions have been scaled by 0.1 to facilitate comparability with the conditional revenue functions.
Source: Own calculations based on CPS-ASEC (2019)

Figure L27: Cond. revenue functions, deciles (2019), low elasticity

(a) Secondary earners



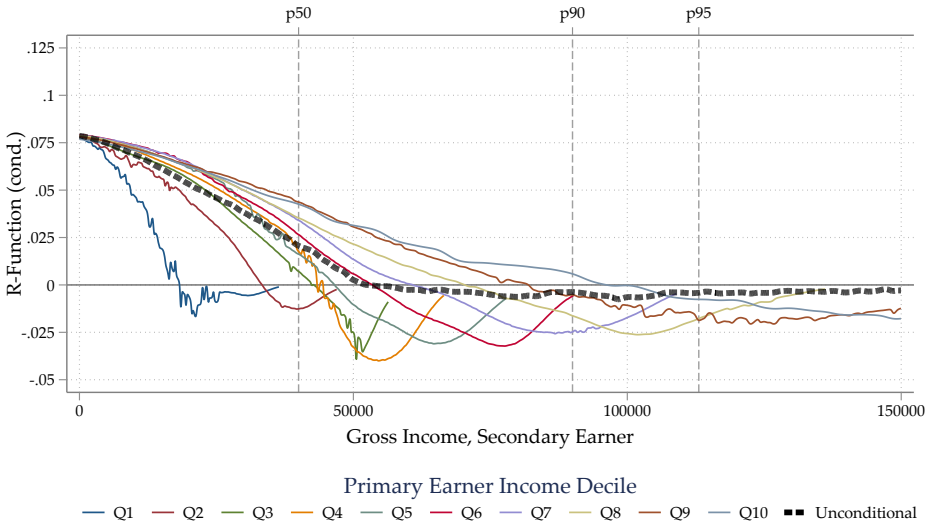
(b) Primary earners



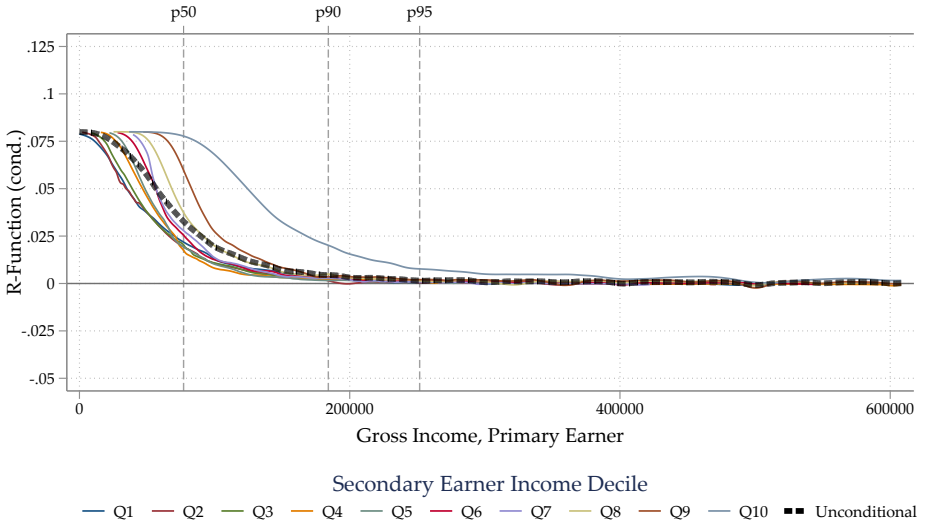
Note: This figure shows conditional revenue functions for secondary (primary) earners in married dual earner couples conditional on primary (secondary) earnings income deciles in Panel a (b) as of 2019. All estimates are based on tax units with non-negative gross income in which spouses are of age 25-55. Gross income contains wage, farm, retirement, and business income, as well as income from dividends, interest, and rent. We take behavioral responses at the intensive and extensive margin into account. We assume intensive margin elasticities of 0.15 (0.35) for primary (secondary) earners and an extensive margin parameter of 0.2. The figure also displays modified unconditional revenue functions for secondary and primary earners where unconditional revenue functions have been scaled by 0.1 to facilitate comparability with the conditional revenue functions. Results for the baseline elasticity scenario are shown in Figure 6 of the main text.
Source: Own calculations based on CPS-ASEC (2019)

Figure L28: Cond. revenue functions, deciles (2019), high elasticity

(a) Secondary earners



(b) Primary earners



Note: This figure shows conditional revenue functions for secondary (primary) earners in married dual earner couples conditional on primary (secondary) earnings income deciles in Panel a (b) as of 2019. All estimates are based on tax units with non-negative gross income in which spouses are of age 25-55. Gross income contains wage, farm, retirement, and business income, as well as income from dividends, interest, and rent. We take behavioral responses at the intensive and extensive margin into account. We assume intensive margin elasticities of 0.5 (1.5) for primary (secondary) earners and an extensive margin parameter of 0.2. The figure also displays modified unconditional revenue functions for secondary and primary earners where unconditional revenue functions have been scaled by 0.1 to facilitate comparability with the conditional revenue functions. Results for the baseline elasticity scenario are shown in Figure 6 of the main text.
Source: Own calculations based on CPS-ASEC (2019)

L.3.4 Alternative conditioning brackets

In the main text, we construct conditioning brackets by looking at deciles of the primary and secondary earner income distribution. Our framework, however, is flexible enough to capture any construction of conditioning brackets. Empirically, we are limited by data constraints, since we need enough observations to estimate conditional distributions. If we condition on wider brackets, we receive more precise results; however we potentially also average out more inefficiencies. If we condition on narrower brackets, we require less averaging, but also have less observations to estimate distributions. In this section, we provide results for wider brackets (quintiles in Figure L29), and more narrow brackets (vigintiles in Figure L30).

An alternative way of conditioning on primary or secondary earnings is to select brackets based on the statutory income tax schedule. Table L4 displays the corresponding income thresholds and the share of primary and secondary earners with incomes in the respective tax bracket. Figure L31 displays the estimated conditional revenue functions for secondary (primary) earners conditional on primary (secondary) earnings being in one of the seven income brackets in 2019. In contrast to the conditioning based on deciles, revenue functions for different tax brackets do not have a common intercept with the vertical axis, because the share of secondary earners in a particular income range is no longer the same. However, conditional revenue functions again show inefficiently high marginal tax rates for some "middle tax bracket" income levels (see Figure L31a).

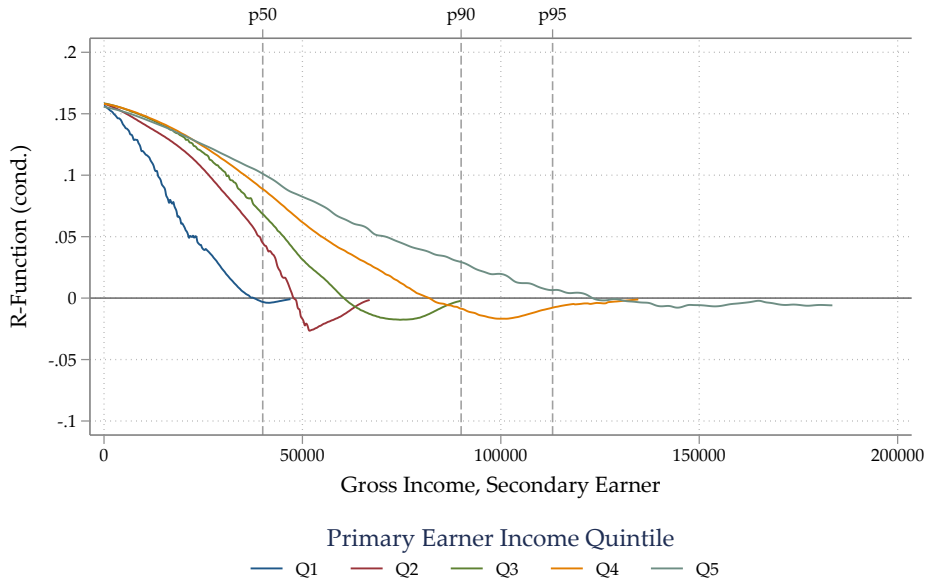
Table L4: Taxable income brackets (2019)

Bracket	Upper threshold	MTR	Share prim. earner	Share sec. earner
1	19400	10	1.83%	20.41%
2	78950	12	48.89%	65.41%
3	168400	22	36.83%	12.75%
4	321450	24	9.46%	1.29%
5	408200	32	1.01%	0.02%
6	612350	35	1.01%	0.05%
7	.	37	0.98%	0.07%

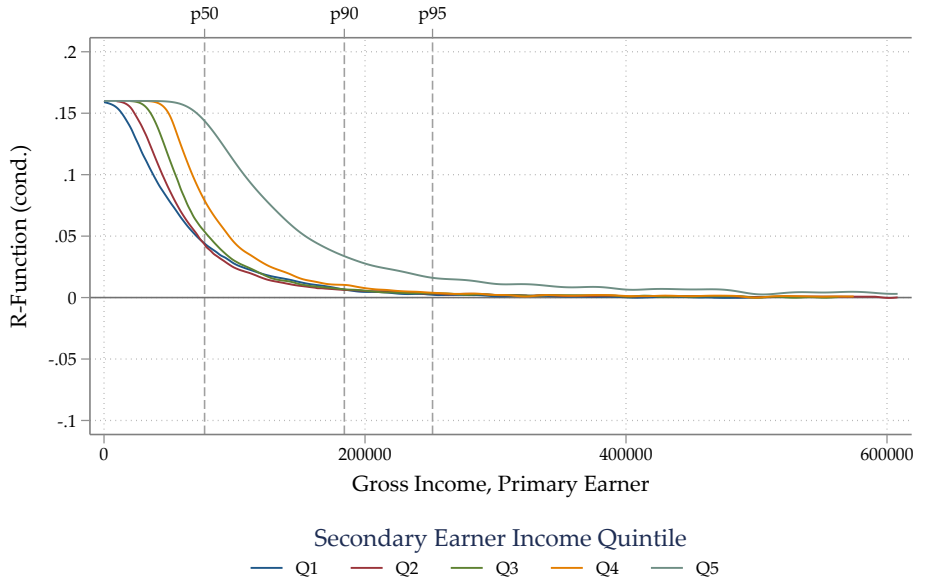
Note: This table shows the upper thresholds (column 2) and marginal tax rates (column 3) for the statutory tax schedule for married couples filing jointly as of 2019. All threshold values refer to taxable income. Column 5 (6) displays the share of primary (secondary) earners whose primary (secondary) earnings lie in a respective bracket.
Source: Own calculations based on CPS-ASEC (2019) and U.S. Federal Individual Income Tax Rates and Brackets (2019)

Figure L29: Conditional revenue functions, quintiles (2019)

(a) Secondary earners



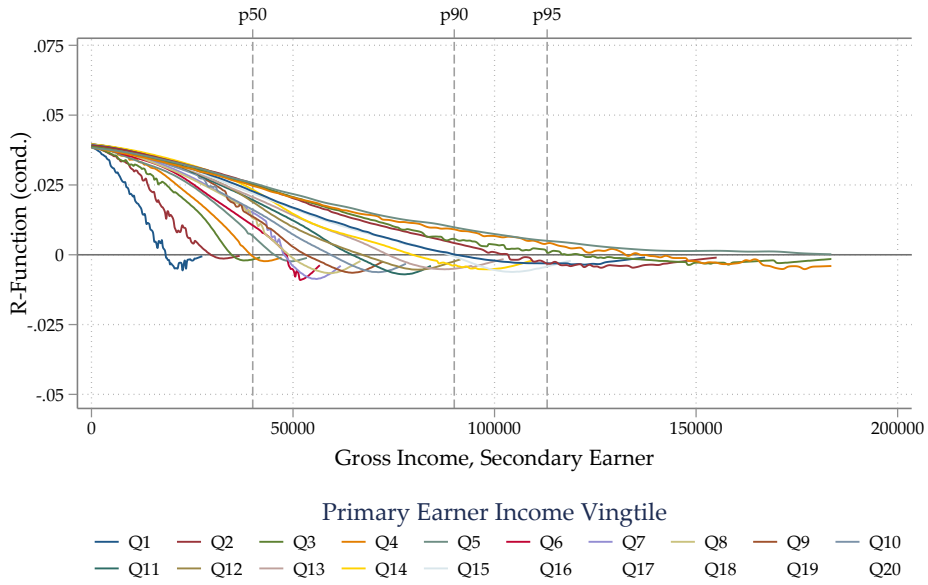
(b) Primary earners



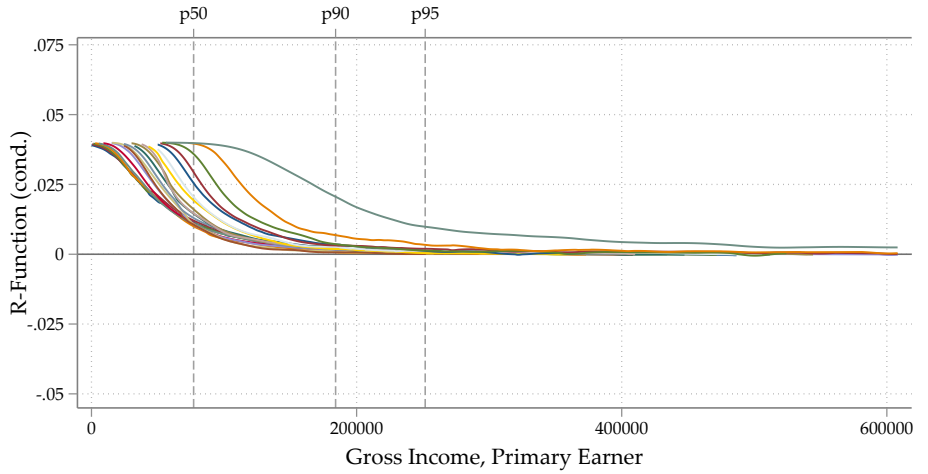
Note: This figure shows conditional revenue functions for secondary (primary) earners in married dual earner couples conditional on primary (secondary) earnings income quintiles in Panel a (b) as of 2019. All estimates are based on tax units with non-negative gross income in which spouses are of age 25-55. Gross income contains wage, farm, retirement, and business income, as well as income from dividends, interest, and rent. We take behavioral responses at the intensive and extensive margin into account. We assume intensive margin elasticities of 0.25 (0.75) for primary (secondary) earners and an extensive margin parameter of 0.2.
Source: Own calculations based on CPS-ASEC (2019)

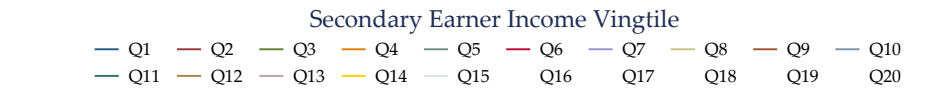
Figure L30: Conditional revenue functions, vigintiles (2019)

(a) Secondary earners



(b) Primary earners

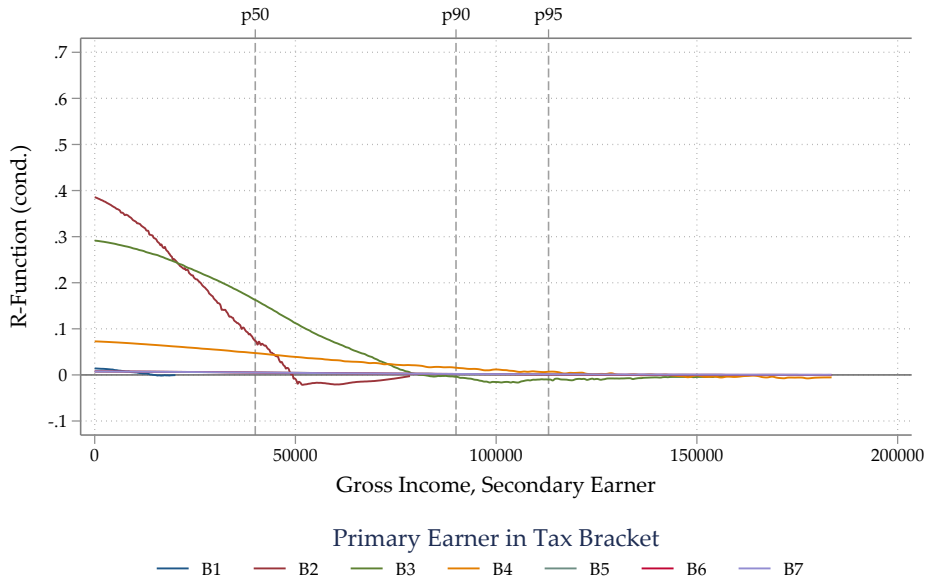




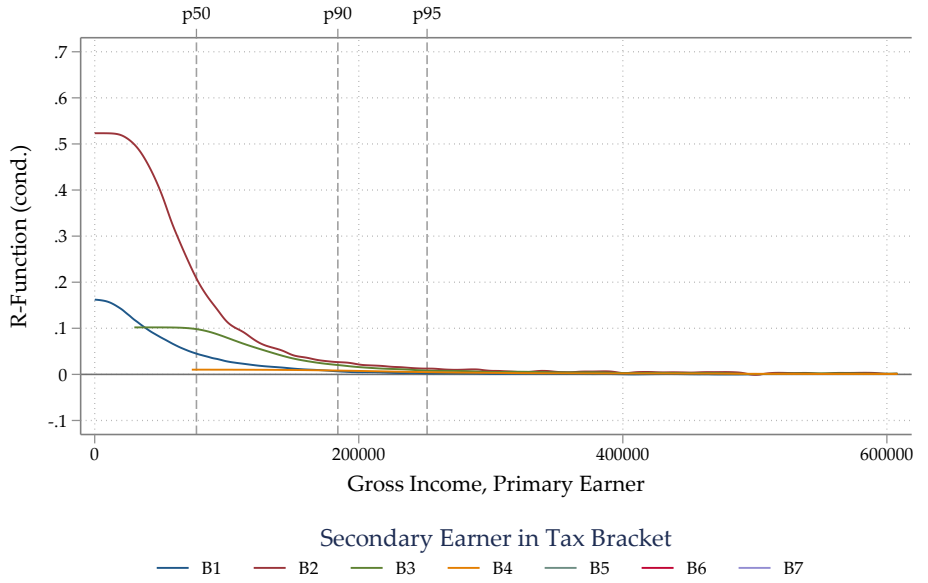
Note: This figure shows conditional revenue functions for secondary (primary) earners in married dual earner couples conditional on primary (secondary) earnings income vigintiles in Panel a (b) as of 2019. All estimates are based on tax units with non-negative gross income in which spouses are of age 25-55. Gross income contains wage, farm, retirement, and business income, as well as income from dividends, interest, and rent. We take behavioral responses at the intensive and extensive margin into account. We assume intensive margin elasticities of 0.25 (0.75) for primary (secondary) earners and an extensive margin parameter of 0.2.
Source: Own calculations based on CPS-ASEC (2019)

Figure L31: Conditional revenue functions, tax brackets (2019)

(a) Secondary earners



(b) Primary earners



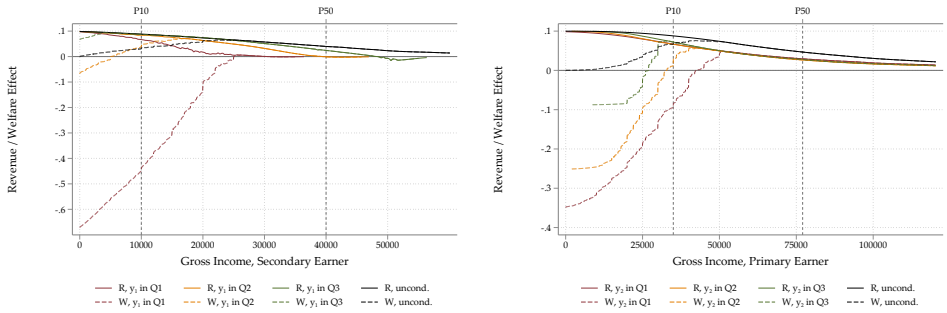
Note: This figure shows conditional revenue functions for secondary (primary) earners in married dual earner couples conditional on primary (secondary) earnings in specific income tax brackets in Panel a (b) as of 2019. All estimates are based on tax units with non-negative gross income in which spouses are of age 25-55. Gross income contains wage, farm, retirement, and business income, as well as income from dividends, interest, and rent. We take behavioral responses at the intensive and extensive margin into account. We assume intensive margin elasticities of 0.25 (0.75) for primary (secondary) earners and an extensive margin parameter of 0.2.
Source: Own calculations based on CPS-ASEC (2019)

L.3.5 Welfare effects of reforms at the bottom

In Section 6.1 of the main text, we also estimate welfare effects of tax reforms at the bottom of the income distribution. This section presents robustness tests on the exclusion of extensive margin responses (Figure L32), details on the shape of average welfare weights (Figure L33) and the choice of elasticities (Figure L34).

Figure L32: Cond. reforms at bottom (2019), no extensive margin

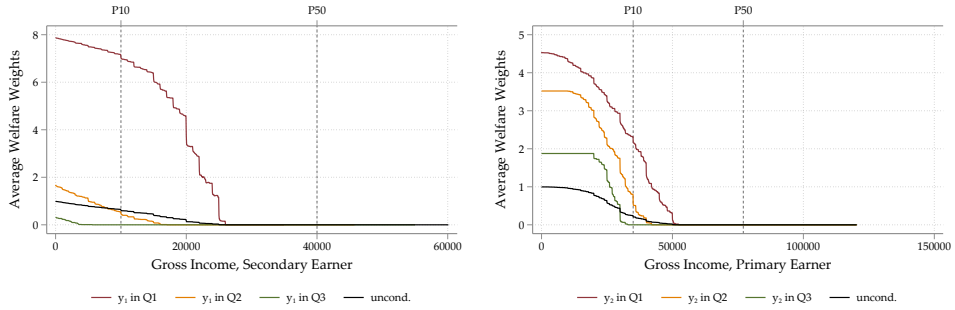
(a) Secondary earnings (b) Primary earnings



Note: This figure replicates Figure 7 without extensive margin responses. The figure shows the revenue and welfare effects for secondary (primary) earners in married dual earner couples conditional on primary (secondary) earnings income deciles at the bottom of the income distribution in Panel a (b) as of 2019. We assume intensive margin elasticities of 0.25 (0.75) for primary (secondary) earners. The figure displays the revenue effect in terms of conditional revenue functions (solid lines) and the welfare effect based on near-Rawlsian welfare weights (dashed lines).
Source: Own calculations based on CPS-ASEC (2019)

Figure L33: Cond. reforms at bottom (2019), welfare weights

(a) Secondary earnings (b) Primary earnings

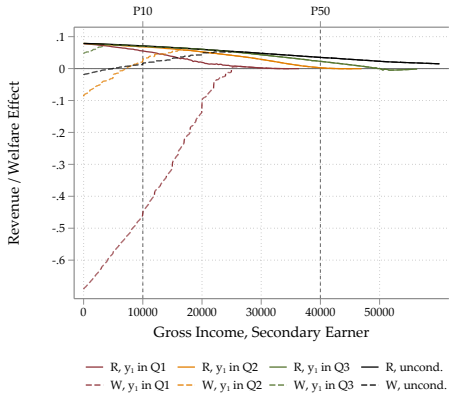


Note: This figure shows average welfare weights used for the calculation of near-Rawlsian welfare effects in Figure 7. Weights for secondary (primary) earners above a specific value conditional on primary (secondary) earnings income deciles at the bottom are shown in Panel a (b). Black lines show unconditional welfare weights. See Table L5 for the exact specification of welfare weights.
Source: Own calculations based on CPS-ASEC (2019)

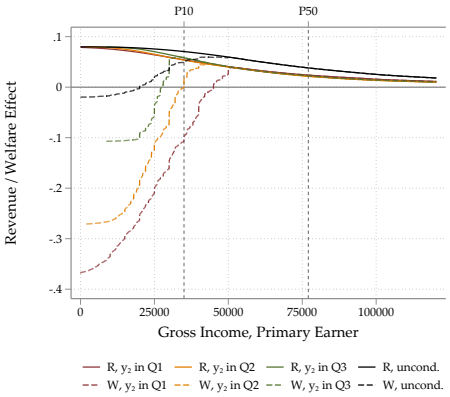
Figure L34: Cond. reforms at bottom (2019), different elasticities

Low elasticity scenario

(a) Secondary earnings

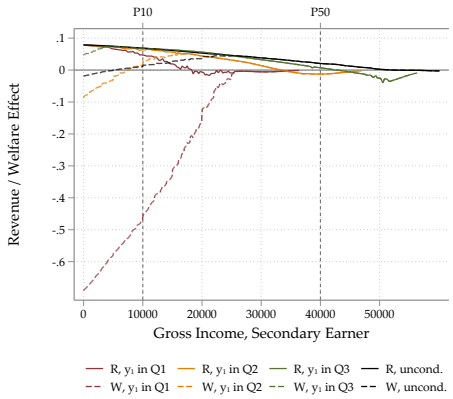


(b) Primary earnings

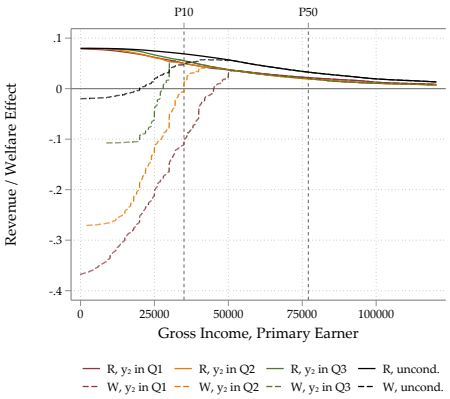


High elasticity scenario

(c) Secondary earnings



(d) Primary earnings



Note: This figure replicates Figure 7 under the low elasticity scenario from Table L2. The figure shows the revenue and welfare effects for secondary (primary) earners in married dual earner couples conditional on primary (secondary) earnings income deciles at the bottom of the income distribution in Panel a (b) as of 2019. All estimates are based on tax units with non-negative gross income in which spouses are of age 25-55. Gross income contains wage, farm, retirement, and business income, as well as income from dividends, interest, and rent. We take behavioral responses at the intensive and extensive margin into account. In the low elasticity scenario, we assume intensive margin elasticities of 0.15 (0.35) for primary (secondary) earners. In the high elasticity scenario, we assume intensive margin elasticities of 0.5 (1.5) for primary (secondary) earners. The extensive margin scaling factor is held constant at 0.9. The figure displays the revenue effect in terms of conditional revenue functions (solid lines) and the welfare effect based on near-Rawlsian welfare weights according to Table L5 (dashed lines). Black lines indicate unconditional revenue and welfare functions.
Source: Own calculations based on CPS-ASEC (2019)

L.4 Reducing jointness

In Section 6.2 of the main text, we estimate the welfare effects of jointness-reducing reforms at different combinations of primary and secondary earnings. Here, we provide detailed information on the construction of empirical welfare assessments (Section L.4.1), and show supplementary results and robustness checks (Section L.4.2).

L.4.1 From theory to data

According to Equation (27), the welfare effect of a local jointness change is given by

$$W_s(0, h^j) \quad = \quad \tau_{12}^j \int_{B(y,\ell)} \mathbf{W}^J(y) dy. \tag{145}$$

For a bracket that is small enough, we can approximate this term as

$$\begin{aligned} W_s(0, h^j) \quad &= \quad \tau_{12}^j \int_{B(y,\ell)} \mathbf{W}^J(y) dy \\ &= \tau_{12}^j \;\; l_1 \;\; l_2 \;\; \mathbf{W}^J(y) \\ &= \tau_{12}^j \;\; l_1 \;\; l_2 \;\; \left(\frac{\mathbf{R}^J(y)}{1 - \Delta_\delta^0} - \int_{y_1}^{\bar{y}_1} \int_{y_2}^{\bar{y}_2} \bar{g}(s_1, s_2) f(s_1, s_2) ds_2 ds_1 \right) \end{aligned} \tag{146}$$

Under $u(c) = c$ and iso-elastic effort costs, there are no income effects, which implies that $\Delta_\delta^0 = 0$. Thus,

$$\begin{aligned} W_s(0, h^j) \quad &= \tau_{12}^j \;\; l_1 \;\; l_2 \;\; \left(\int_{y_2}^{\bar{y}_2} \bar{\mathcal{R}}_1(y_1, s_2) ds_2 \; + \int_{y_1}^{\bar{y}_1} \bar{\mathcal{R}}_2(s_1, y_2) ds_1 \quad + \int_{y_1}^{\bar{y}_1} \int_{y_2}^{\bar{y}_2} \bar{\mathcal{R}}_b(s_1, s_2) ds_2 ds_1 \right. \\ &\quad \left. - \int_{y_1}^{\bar{y}_1} \int_{y_2}^{\bar{y}_2} \bar{g}(s_1, s_2) f(s_1, s_2) ds_2 ds_1 \right) \end{aligned} \tag{147}$$

Using the base ingredients from Equation (139),

$$\begin{aligned}\int_{y_2}^{\bar{y}_2} \bar{\mathcal{R}}_1(y_1, s_2) ds_2 &= \int_{y_2}^{\bar{y}_2} -f(y_1, s_2)\, \varepsilon_1\, y_1\, \frac{T_1(y_1, s_2)}{1 - T_1(y_1, s_2)} ds_2 \\ &= \underbrace{s(s_2 \geq y_2)}_{\text{Share}} \Bigg(\underbrace{- \mathbb{E}_{y_1}\left[\varepsilon_1\, y_1\, \frac{T_1(y_1, s_2)}{1 - T_1(y_1, s_2)} | s_2 > y_2\right] f^{y_1}(y_1 \mid s_2 \geq y_2)}_{\text{Intensive}} \Bigg)\end{aligned} \tag{148}$$

The term $\int_{y_1}^{\bar{y}_1} \bar{\mathcal{R}}_2(s_1, y_2) ds_1$ describing mechanical effects is estimated as

$$\begin{aligned}\int_{y_1}^{\bar{y}_1}\int_{y_2}^{\bar{y}_2} \bar{\mathcal{R}}_b(s_1, s_2) ds_2 ds_1 &= \int_{y_1}^{\bar{y}_1}\int_{y_2}^{\bar{y}_2} f^e(s_1, s_2)\big(1 - L_E\big) ds_2 ds_1. \\ &= \underbrace{s(s_1 \geq y_1)}_{\text{Share}} \underbrace{\Big(1 - F^{s_2}\big(s_2 \mid s_1 \geq y_1\big)\Big)}_{\text{Cond. CDF}} \underbrace{(1 - L_E)}_{\text{Extensive}} \\ &= \underbrace{s(s_2 \geq y_2)}_{\text{Share}} \underbrace{\Big(1 - F^{s_1}\big(s_1 \mid s_2 \geq y_2\big)\Big)}_{\text{Cond. CDF}} \underbrace{(1 - L_E)}_{\text{Extensive}} \\ &= \underbrace{s(s_1 \geq y_1, s_2 \geq y_2)}_{\text{Share}} \underbrace{(1 - L_E)}_{\text{Extensive}}\end{aligned} \tag{149}$$

Finally, the term governing the welfare costs is estimated as

$$\int_{y_1}^{\bar{y}_1}\int_{y_2}^{\bar{y}_2} \bar{g}(s_1, s_2) f(s_1, s_2) ds_2 ds_1 = \underbrace{s(s_1 \geq y_1, s_2 \geq y_2)}_{\text{Share}} \underbrace{\mathbb{E}[g(s_1, s_2) | s_1 \geq y_1, s_2 \geq y_2]}_{\text{Weights}} \tag{150}$$

Combining all equations, the empirical welfare effect reads

$$\begin{aligned} W_s(0, h^j) \quad = \quad & \underbrace{\tau_{12}^j \;\; l_1 \;\; l_2 \;\; s(s_2 \geq y_2) \Bigg(- \mathbb{E}_{y_1}\left[\varepsilon_1\, y_1\, \frac{T_1(y_1, s_2)}{1 - T_1(y_1, s_2)} | s_2 > y_2\right] f^{y_1}(y_1 \mid y_2 \geq s_2) \Bigg)}_{\text{Intensive (P)}} \\ + \quad & \underbrace{\tau_{12}^j \;\; l_1 \;\; l_2 \;\; s(s_1 \geq y_1) \Bigg(- \mathbb{E}_{y_2}\left[\varepsilon_2\, y_2\, \frac{T_1(s_1, y_2)}{1 - T_1(s_1, y_2)} | s_1 > y_1\right] f^{y_2}(y_2 \mid s_1 \geq y_1) \Bigg)}_{\text{Intensive (S)}} \\ + \quad & \underbrace{\tau_{12}^j \;\; l_1 \;\; l_2 \;\; s(s_1 \geq y_1, s_2 \geq y_2)}_{\text{Mechanical}} - \underbrace{\tau_{12}^j \;\; l_1 \;\; l_2 \;\; s(s_1 \geq y_1, s_2 \geq y_2) \;\; L_E}_{\text{Extensive}} \\ - \quad & \underbrace{\tau_{12}^j \;\; l_1 \;\; l_2 \;\; s(s_1 \geq y_1, s_2 \geq y_2) \;\; \mathbb{E}[g(s_1, s_2) | s_1 \geq y_1, s_2 \geq y_2]}_{\text{Weights}} \end{aligned} \tag{151}$$

where the first two terms are revenue effects coming behavioral responses triggered by the change in marginal tax rates for primary and secondary earners. The third term is the pure mechanical effect of the reform without extensive margin responses

while the fourth part represents the loss of mechanical revenue due to extensive margin responses. The final term captures the adjustment required by specific welfare objectives.

Specifications. For consistency reasons, assumptions about behavioral responses follow the previous section (see Table L2). We use three different specifications of welfare weights displayed in Table L5. Reforms are carried out at bracket combinations defined by the primary earnings deciles and the respective conditional secondary earnings deciles (see Table L6).

Table L5: Welfare weights

Welfare Weight Type	Welfare Weight
Equal Weights	$g(y_1, y_2) = 1$
Near-Rawlsian Weights	$g(y_1, y_2) = \begin{cases} 1, & \text{for } y_1 + y_2 \leq \mathrm{P}_{10} \\ 0, & \text{for } y_1 + y_2 \geq \mathrm{P}_{10} \end{cases}$
Feminist Weights	$g(y_1, y_2) = \begin{cases} \frac{y_2}{y_1+y_2}, & \text{for } y_2 = y_{\text{woman}} \\ \frac{y_1}{y_1+y_2}, & \text{for } y_1 = y_{\text{woman}} \end{cases}$

Note: This table shows the specifications of different welfare weights used in the empirical analysis. For near-Rawlsian weights, P_{10} refers to the tenth percentile of the joint income distribution of dual earner couples. Welfare weights are normalized to mean one.

Table L6: Reforms on jointness, selected reform brackets (2019)

Primary earnings bracket	$\underline{y}_1$	Secondary earnings bracket, $\underline{y}_2$									
		$B_2(1)$	$B_2(2)$	$B_2(3)$	$B_2(4)$	$B_2(5)$	$B_2(6)$	$B_2(7)$	$B_2(8)$	$B_2(9)$	$B_2(10)$
$B_1(1)$	1	1	3200	7488	10167	14001	16001	19001	20010	24001	27750
$B_1(2)$	35001	1	7500	15002	20001	24000	28002	30115	34001	36001	40001
$B_1(3)$	46003	1	8484	18006	25001	30001	34011	38001	41802	45603	50001
$B_1(4)$	55625	1	12001	20001	27002	33501	40001	45001	49792	52002	56004
$B_1(5)$	65255	1	14288	23150	30086	36040	42202	49550	53700	60006	65353
$B_1(6)$	77060	1	16102	26523	35012	42005	50001	55271	61100	70005	76401
$B_1(7)$	91036	1	18200	30005	39200	46225	52800	60124	70010	79501	88060
$B_1(8)$	109000	1	16200	28307	37757	47000	55028	65125	76150	90001	101051
$B_1(9)$	135031	1	16001	30005	42510	55001	65218	77252	91088	107345	128001
$B_1(10)$	184202	1	15106	30014	45200	59100	73500	92301	114400	140223	186300

Note: This table shows the lower bracket thresholds used for the analysis of different jointness-reducing reforms. Primary earner brackets (vertical) are chosen based on deciles of the primary earner income distribution. Secondary earner brackets (horizontal) are chosen based on the secondary earner deciles conditional on the respective primary earner decile. The red (blue) cells indicate the bracket combination, for which reforms are discussed in detail (Figure L36 and Figure L37).
Source: Own calculations based on CPS-ASEC (2019)

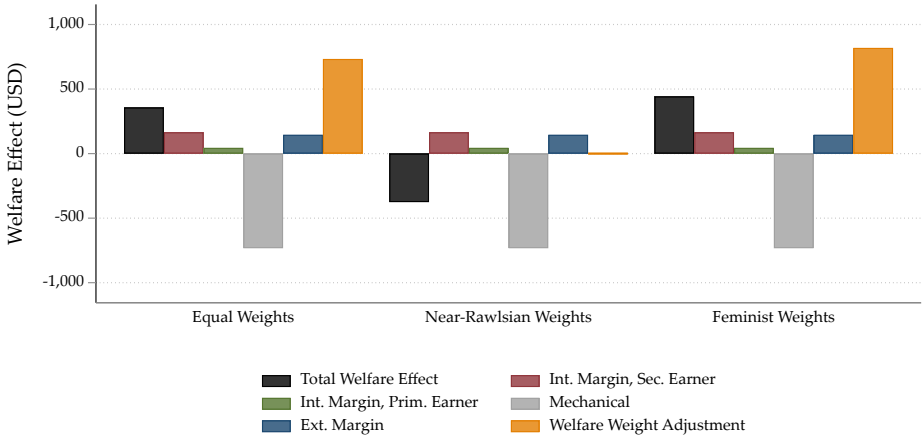
L.4.2 Supplementary graphs and robustness

In the main text, we provide a high-level overview of the welfare effects at different combinations of primary and secondary earnings (Figure 8). Here, we provide

supplementary graphs on the construction and decomposition of the welfare effect.

Figure L35 illustrates the welfare effect for a reform that reduces jointness in $B_1 \times B_2$, where B_1 starts at the median level of primary earnings and y_2 starts at the median of secondary earnings conditional on $y_1 \in B_1$. It provides a decomposition into the various components that appear on the right hand side of Equation (151). The left hand side corresponds to the black bar. The Figure shows that the reform is rejected by a near-Rawlsian welfare function and approved both by a Feminist one and by one with equal weights.

Figure L35: Welfare effect, reducing jointness at the median (2019)

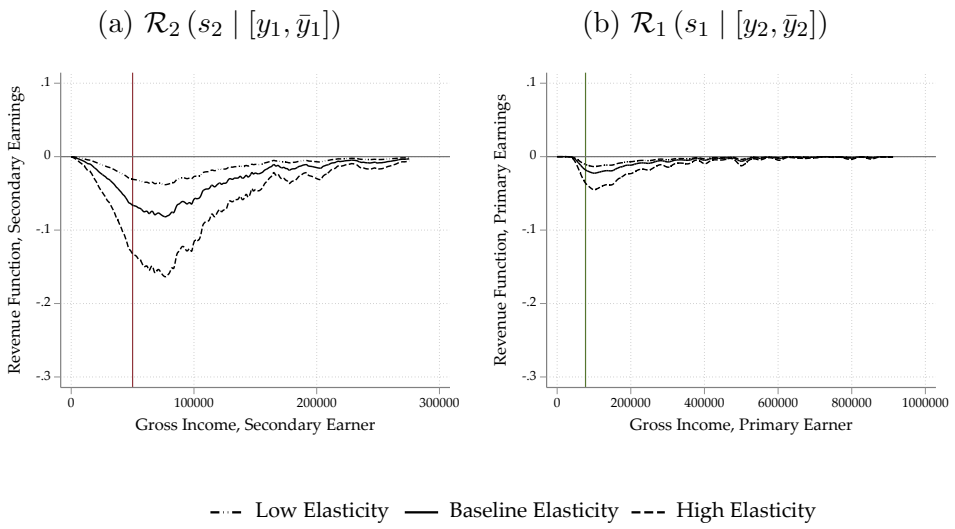


Note: This figure visualizes the welfare effects implied by a reduction in jointness as of 2019 for $(s_1, s_2) \in B_1 \times B_2$, where B_1 denotes the 6th decile of primary earnings and B_2 the 6th decile of secondary earnings conditional on $s_1 \in B_1$. The aggregate welfare effect and its components are shown for three welfare weights (Table L5), based on bracket lengths $\ell_1 = \ell_2 = 500$ and $\tau = -0.01$. Intensive margin responses follow the baseline elasticity scenario (Table L2), and extensive margin responses are also included with $L_E = 0.2$. The black bar indicates the total welfare effect; red (green) bars show the revenue change from intensive responses of secondary (primary) earnings; the gray bar depicts the mechanical effect; the blue bar the counteracting extensive margin effect; and the orange bar the welfare-weight adjustment corresponding to the final term in equation (151). The conditional revenue functions underlying the intensive margin responses are illustrated in Figure L36. For further details, see Appendix L.4.
Source: Own calculations based on CPS-ASEC (2019).

Figure L36 provides supplementary details on the empirical conditional revenue functions that are used to estimate the welfare effect for a reduction of jointness around the median in Figure L35.

Note that the revenue effects coming from the respective changes in secondary and primary earner marginal tax rates can vary substantially in magnitude. For instance, when marginal tax rates are inefficiently high, e.g. under strong behavioral responses and for relatively rich secondary earners (see Figure L37), the (conditional) cut of secondary earner marginal tax rates triggers a revenue gain

Figure L36: Welfare effect, reducing jointness at $B_1(6)$, $B_2(6)$

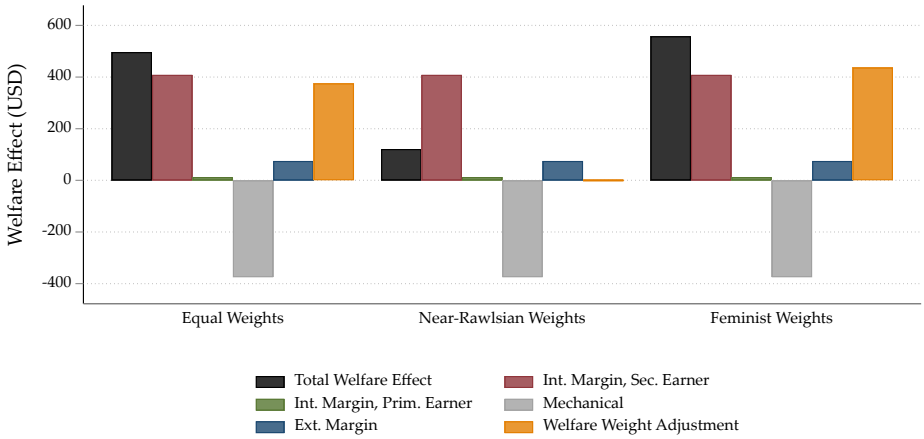


Note: This figure visualizes the estimation of the two revenue effects from intensive margin responses required for estimating the welfare effect implied by a reduction in jointness for $(s_1, s_2) \in B_1 \times B_2$, where B_1 is the 6th-decile of the distribution of primary earnings and B_2 is the sixth decile in the distribution of secondary earnings conditional on $s_1 \in B_1$ (see Figure L35). Figure L36b shows the relevant conditional revenue function describing the effect of increasing marginal tax rates for primary earners above a certain threshold conditional on secondary earnings being above the bracket threshold y_2 of the bracket $B_2(6)$. The solid red (green) lines indicate the lower bracket thresholds of secondary (primary) earnings that are relevant for the welfare evaluation, i.e. y_2 (y_1). Both revenue functions distinguish between different intensive margin elasticity scenarios and include extensive margin behavioral responses.
Source: Own calculations based on CPS-ASEC (2019)

that outweighs the mechanical revenue loss. In such cases, even near-Rawlsian welfare agrees with decreases in jointness, and the conflict between different welfare objectives disappears.

Figure L38 replicates the results on the welfare implications of changing jointness at different bracket combinations and illustrates, how results change under different elasticity scenarios. Figure L39 provides the results without extensive margin responses.

Figure L37: Welfare effect, reducing jointness at $B_1(6)$, $B_2(10)$

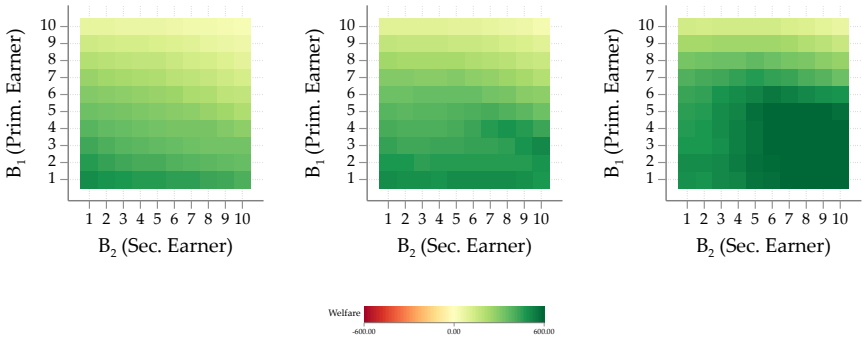


Note: This figure visualizes the decomposition of welfare effects implied by a reduction in jointness as of 2019 for $(s_1, s_2) \in B_1 \times B_2$, where B_1 denotes the 6th decile of primary earnings and B_2 the 10th decile of secondary earnings conditional on $s_1 \in B_1$. The aggregate welfare effect and its components are shown for three welfare weights (Table L5), based on bracket lengths $\ell_1 = \ell_2 = 500$ and $\tau = -0.01$. Intensive margin responses follow the high elasticity scenario (Table L2), and extensive margin responses are also included with $L_E = 0.2$. The black bar indicates the total welfare effect; red (green) bars show the revenue change from intensive responses of secondary (primary) earnings; the gray bar depicts the mechanical effect; the blue bar the counteracting extensive margin effect; and the orange bar the welfare-weight adjustment corresponding to the final term in equation (151).
Source: Own calculations based on CPS-ASEC (2019)

Figure L38: Decreasing jointness, varying elasticities

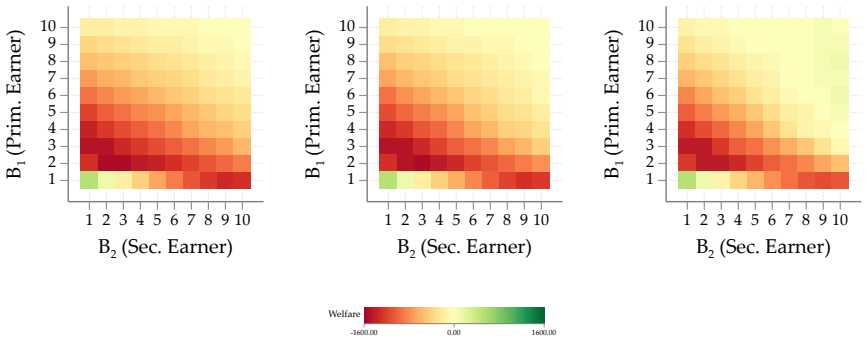
Equal weights

(a) Low elasticity (b) Baseline elasticity (c) High elasticity



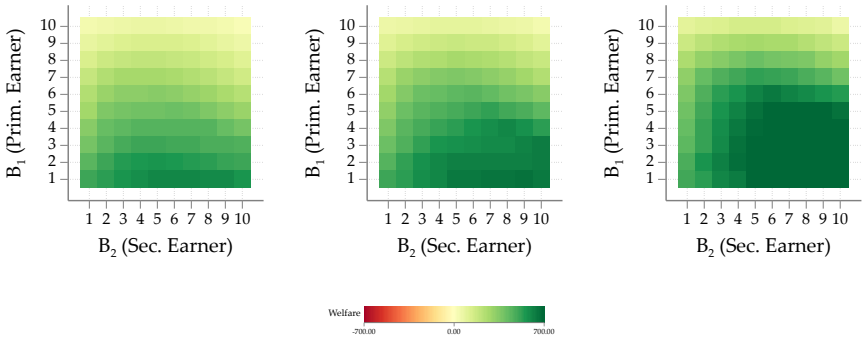
Near-Rawlsian weights

(d) Low Elasticity (e) Baseline Elasticity (f) High Elasticity



Feminist weights

(g) Low elasticity (h) Baseline elasticity (i) High elasticity

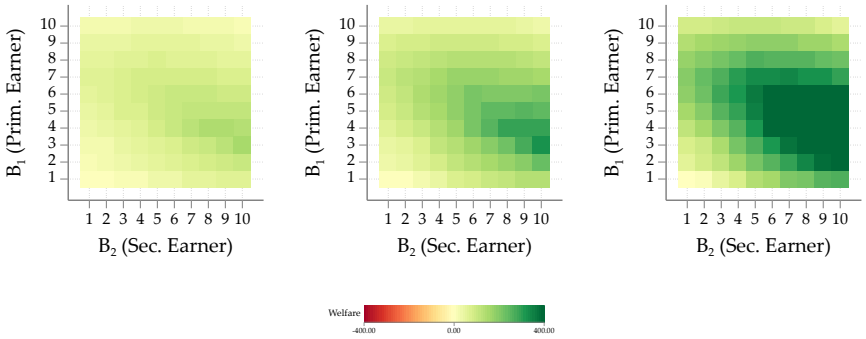


Note: This figure displays the welfare implications of decreasing jointness by decreasing secondary and primary marginal tax rates conditional on particular brackets of primary earnings and secondary earnings. Bracket thresholds are shown in Table L6. The reform applies to a bracket of length $l(B_1) = l(B_2) = 500$ and has the magnitude $\tau = -0.01$. The figure distinguishes between three different forms of welfare weights (see Table L5) and three elasticity scenarios (see Table L2). All results are shown including extensive margin responses. Results without extensive margin responses are shown in Table L39.
Source: Own calculations based on CPS-ASEC (2019)

Figure L39: Decreasing jointness, no extensive margin

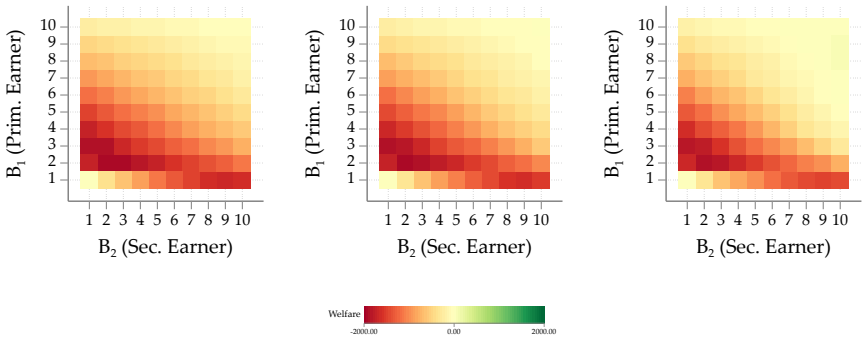
Equal weights

(a) Low elasticity (b) Baseline elasticity (c) High elasticity



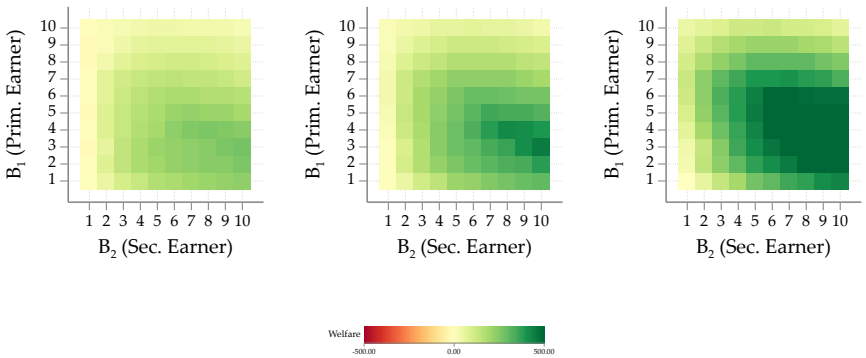
Near-Rawlsian weights

(d) Low elasticity (e) Baseline elasticity (f) High elasticity



Feminist weights

(g) Low elasticity (h) Baseline elasticity (i) High elasticity



Note: This figure displays the welfare implications of decreasing jointness by decreasing secondary and primary marginal tax rates conditional on particular brackets of primary earnings and secondary earnings. Bracket thresholds are shown in Table L6. The reform applies to a bracket of length $l(B_1) = l(B_2) = 500$ and has the magnitude $\tau = -0.01$. The figure distinguishes between three different forms of welfare weights (see Table L5) and three elasticity scenarios (see Table L2). All results are shown excluding extensive margin responses. Results with extensive margin responses are shown in Figure L38.
Source: Own calculations based on CPS-ASEC (2019)

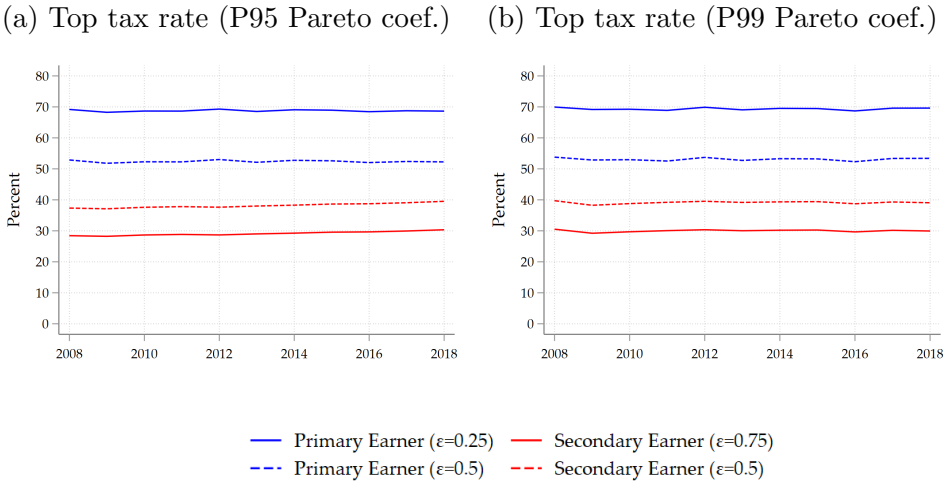
L.5 Optimal top tax rates

In Section 6.3 of the main text, we use tabular income tax information on all married couples for the estimation of Pareto coefficients and associated optimal tax rates in the most recent available year. In the following, we provide insights on the stability of optimal tax rates over time (Section L.5.1), heterogeneity analyses that estimate Pareto coefficients and optimal tax rates separately for single earner and dual earner couples (Section L.5.2), and robustness tests that estimate Pareto coefficients from CPS data (Section L.5.3).

L.5.1 Optimal tax rates over time

Figure L40 presents optimal tax rates based on Pareto coefficients that have been calculated at the 95th and 99th percentile for all available years of the tabulated income tax return data. Since the distributional properties did not change over this relatively recent time horizon, optimal top tax rates stay rather constant.

Figure L40: Optimal top tax rates over time



Note: This figure displays optimal top tax rates for different values of the Frisco elasticity based on the Pareto coefficients estimated from tabulated data on wages on W2 forms for joint return taxpayers with wage income.
Source: Own calculations based on SOI Tax Stats - Individual Information Return Form W-2 (2018)

L.5.2 Heterogeneity analyses

Tables L7 (L8) display Pareto coefficients and respective optimal tax rates for single earner couples (dual earner couples with relatively equal income shares). The tables

provide two insights. First, the Pareto coefficient for single earner couples (primary earners) is much lower than for dual earner couples (primary earner and secondary earner) pushing towards higher optimal tax rates for single earner couples compared to dual earner couples. Second, within dual earner couples, the optimal tax rate gap between primary and secondary earners almost vanishes, since Pareto coefficients are very similar.

Table L7: Pareto coefficients and optimal tax rates, single earner couples (2018)

	P95		P99	
	Primary	Secondary	Primary	Secondary
Panel A: Pareto coefficients				
Wages on W2 form	1.65	-	1.61	-
Panel B: Optimal tax rates				
Elasticity = .25	71%	-	71%	-
Elasticity = .5	55%	-	55%	-
Elasticity = .75	45%	-	45%	-

Note: This table shows Pareto coefficients and optimal top tax rates. Panel A displays Pareto coefficients for primary and secondary earners based on a generalized Pareto interpolation using tabulated data on wages on W2 forms for joint return single earner couples with wage income. We distinguish between interpolations at the 95th and the 99th percentile. Panel B displays optimal top tax rates associated with these Pareto coefficients. We distinguish between different elasticities.
Source: Own calculations based on SOI Tax Stats - Individual Information Return Form W-2 Statistics (2018)

Table L8: Pareto coefficients and optimal tax rates, dual earner couples (2018)

	P95		P99	
	Primary	Secondary	Primary	Secondary
Panel A: Pareto coefficients				
Wages on W2 form	2.95	3.07	3.16	3.22
Panel B: Optimal tax rates				
Elasticity = .25	58%	57%	56%	55%
Elasticity = .5	40%	39%	39%	38%
Elasticity = .75	31%	30%	30%	29%

Note: This table shows Pareto coefficients and optimal top tax rates. Panel A displays Pareto coefficients for primary and secondary earners based on a generalized Pareto interpolation using tabulated data on wages on W2 forms for joint return dual earner couples with wage income. Dual earner couples have a secondary earner income share of at least 25 percent. We distinguish between interpolations at the 95th and the 99th percentile. Panel B displays optimal top tax rates associated with these Pareto coefficients. We distinguish between different elasticities.
Source: Own calculations based on SOI Tax Stats - Individual Information Return Form W-2 Statistics (2018)

L.5.3 Robustness

In the main text, we estimate Pareto coefficients with tabular income tax return data that provides high quality at the top of the distribution. However, we can also calculate Pareto coefficients directly using CPS data. For this purpose, we estimate the empirical cumulative distribution function (ECDF) and plot the log-transformed ECDF against the log-transformed income level for specific percentiles (see Figure L41). The slope of the regression lines indicate the Pareto coefficient. In line with main results from tabulated tax return data, the Pareto coefficient for primary earners is larger than the one for secondary earners. This holds irrespective of the type of income used in the estimation and the sample restriction (see Table L9). As shown in Table L10, the optimal tax rate gap between primary and secondary earners is even larger than what is suggested by the tabulated data.

Table L9: Pareto coefficients, CPS (2018)

Sample restr.	Income concept	Top 5%		Top 1%	
		Prim. earner	Sec. earner	Prim. earner	Sec. earner
Baseline	Wages	1.63 (N=933)	3.35 (N=953)	1.37 (N=206)	3.27 (N=194)
Baseline	Wage, Business, Farm Income	1.61 (N=942)	3.33 (N=1022)	1.37 (N=194)	3.33 (N=216)
Baseline	Gross Income	1.64 (N=925)	3.28 (N=933)	1.42 (N=187)	3.28 (N=185)
Alternative	Wages	1.72 (N=1700)	3.25 (N=1743)	1.42 (N=378)	3.21 (N=358)
Alternative	Wage, Business, Farm Income	1.69 (N=1789)	3.13 (N=1871)	1.4 (N=350)	3.04 (N=351)
Alternative	Gross Income	1.74 (N=1705)	3.05 (N=1729)	1.47 (N=348)	3 (N=333)

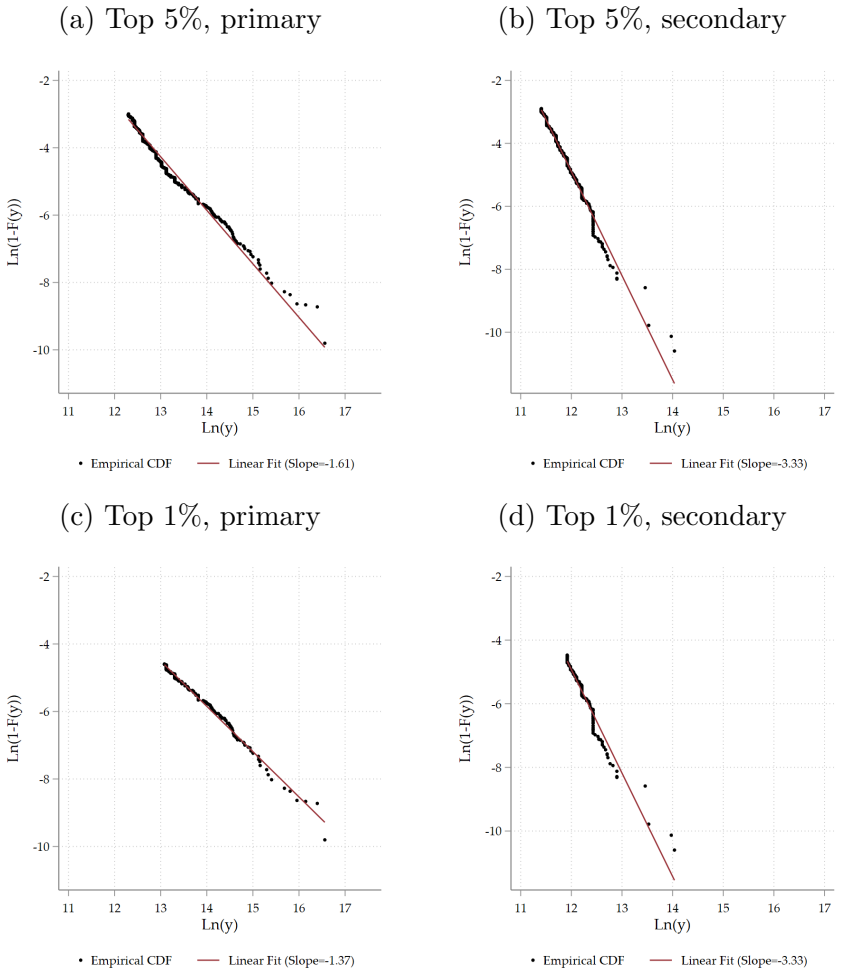
Note: This table shows the Pareto coefficients estimated for different sample restrictions, different income concepts, and at different parts of the income distribution. The baseline sample is restricted to all married couples where both spouses are between 25 and 55 years old. Under an alternative sample restriction, we include all adult married spouses.
Source: Own calculations based on CPS-ASEC (2018)

Table L10: Optimal top tax rates, CPS (2018)

Elasticity	P95		P99	
	Prim. earner	Sec. earner	Prim. earner	Sec. earner
.25	71%	55%	74%	55%
.5	55%	38%	59%	38%
.75	45%	29%	49%	29%

Note: This table shows optimal top tax rates for different values of the Frisch elasticity based on the Pareto coefficients estimated from CPS data under the baseline sample restriction where we focus on all married couples where both spouses are between 25 and 55 years old. Pareto coefficient have been estimated using earnings of primary and secondary earner that include wage, business, and farm income.
Source: Own calculations based on CPS-ASEC (2018)

Figure L41: Pareto coefficient, CPS data



Note: This figure displays the top earnings inequality among primary and secondary earners as of 2018. The sample is restricted to all married couples where both spouses are of age 25-55. The figure shows the shape of the empirical CDF separately for primary and secondary earners in married couples, i.e. $ln(1 - F^e(y))$ on the vertical and $ln(y)$ on the horizontal axis. The figure also shows the linear fit as a red line, the slope of which is the estimated Pareto coefficient. Earnings of primary and secondary earner include wage, business, and farm income.
Source: Own calculations based on CPS-ASEC (2018)